

**IMPLEMENTASI ALGORITMA *LONG SHORT TERM MEMORY (LSTM)*
DALAM MEMPREDIKSI KONSUMSI ENERGI BARU TERBARUKAN**

SKRIPSI

DITA ANGGREANI

211401002



**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA**

MEDAN

2025

**IMPLEMENTASI ALGORITMA *LONG SHORT TERM MEMORY (LSTM)*
DALAM MEMPREDIKSI KONSUMSI ENERGI BARU TERBARUKAN**

SKRIPSI

Diajukan untuk melengkapi tugas dan memenuhi syarat memperoleh ijazah
Sarjana Ilmu Komputer

DITA ANGGREANI

211401002



**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA**

MEDAN

2025

PERSETUJUAN

Judul : IMPLEMENTASI ALGORITMA *LONG SHORT*
TERM MEMORY (LSTM) DALAM MEMPREDIKSI
KONSUMSI ENERGI BARU TERBARUKAN

Kategori : SKRIPSI

Nama : DITA ANGGREANI

Nomor Induk Mahasiswa : 211401002

Program Studi : SARJANA (S-1) ILMU KOMPUTER

Fakultas : ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

Komisi Pembimbing :

Pembimbing II

Pembimbing I

Dr. Eng Ade Candra S.T., M.Kom.
NIP. 197909042009121002

Dewi Sartika Br Ginting, S.Kom., M.Kom.
NIP. 199005042019032023

Diketahui/disetujui oleh
Program Studi S-1 Ilmu Komputer
Ketua

Dr. Amalia, S.T, M.T
NIP. 197812212014042001

PERNYATAAN**IMPLEMENTASI ALGORITMA *LONG SHORT TERM MEMORY (LSTM)*
DALAM MEMPREDIKSI KONSUMSI ENERGI BARU TERBARUKAN****SKRIPSI**

Saya mengakui bahwa skripsi ini adalah hasil penelitian saya sendiri, kecuali beberapa kutipan dan ringkasan yang telah disebutkan sumbernya

Medan, 12 Desember 2024



Dita Anggreani

211401002

PENGHARGAAN

Puji dan syukur kehadiran Allah SWT atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penulisan skripsi ini dengan sebaik-baiknya sebagai syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi S-1 Ilmu Komputer, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara.

Pada kesempatan ini, penulis menyampaikan ungkapan terima kasih dan rasa hormat sebesar-besarnya kepada semua pihak yang telah membantu penulis dalam bentuk doa, dukungan, bimbingan, dan ilmu dalam penyusunan skripsi ini, yaitu:

1. Bapak Dr. Muryanto Amin, S.Sos, M.Si selaku Rektor Universitas Sumatera Utara.
2. Ibu Dr. Maya Silvi Lidya, B.Sc, M.Sc selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.
3. Ibu Dr. Amalia, S.T, M.T selaku Ketua Program Studi S-1 Ilmu Komputer Universitas Sumatera Utara yang telah berdedikasi penuh untuk meluangkan waktu sehingga dapat senantiasa memberikan masukan serta dukungan kepada penulis sepanjang masa studi.
4. Ibu Fuzy Yustika Manik, S.Kom., M.Kom selaku Dosen Pembimbing Akademik yang telah membimbing dan mengarahkan penulis sepanjang masa studi.
5. Ibu Dewi Sartika Br Ginting, S.Kom, M.Kom selaku Dosen Pembimbing I yang telah banyak membimbing dan memberikan saran dalam penyusunan skripsi.
6. Bapak Dr. Eng Ade Candra, S.T., M.Kom selaku Dosen Pembimbing II yang sudah membimbing dan memberikan arahan dalam penyusunan skripsi.
7. Seluruh Bapak/Ibu dosen Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara yang telah berbagi ilmu, wawasan dan pengalaman berharga selama masa studi.
8. Seluruh staf pengajar dan pegawai Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara yang telah membantu dalam proses penyusunan skripsi.

9. Ibunda penulis yang senantiasa menemani penulis pada setiap langkah selama masa studi, selalu memanjatkan doa-doa tulus, kasih sayang dan seluruh pengorbanan yang telah dilalui. Penulis juga mengucapkan banyak terima kasih untuk seluruh keluarga penulis yang menemani dan mendukung penulis selama masa studi dan proses penyusunan skripsi ini.
10. Bapak M. Heru Prasetyo selaku Mentor MSIB Batch 6 PT. Pertamina Internasional EP dan seluruh pegawai divisi Development & Drilling PIEP yang telah berbagi pengalaman dan memberikan semangat kepada penulis.
11. Teman-teman mahasiswa yang telah kebersamai dan mendukung penulis dalam memperluas pemahaman dan wawasan.
12. Semua pihak yang telah membantu saya dalam penelitian ini, baik secara langsung maupun tidak langsung, yang tidak dapat saya sebutkan satu per satu. Terima kasih atas bantuan, waktu, dan kontribusi yang diberikan.

Sebagai penutup, penulis berharap semoga skripsi ini dapat memberikan manfaat bagi para pembaca. Penulis memohon maaf atas segala kekurangan yang ada dalam penyusunan skripsi ini. Kepada seluruh pihak yang telah berjasa dalam proses penyusunan skripsi hingga selesai, penulis ucapkan terima kasih.

Medan, 12 Desember 2024

Penulis,



Dita Anggreani

ABSTRAK

Konsumsi energi global yang terus meningkat mendorong negara-negara di dunia untuk mencari sumber energi alternatif. Dengan menipisnya ketersediaan bahan bakar fosil, energi baru terbarukan menjadi solusi yang dapat dipilih untuk memenuhi kebutuhan energi global. Namun, belum semua negara memiliki perkembangan teknologi yang memadai untuk mengimplementasikan penggunaan energi baru terbarukan sebagai sumber energi. Oleh karenanya dalam penelitian ini dibangun model prediksi konsumsi energi baru terbarukan dengan *deep learning*. Data yang digunakan adalah data konsumsi energi tahunan pada negara berkembang di Asia termasuk diantaranya Bangladesh, Brunei Darussalam, China, Filipina, India, Indonesia, Iran, Kamboja, Laos, Malaysia, Myanmar, Pakistan, Thailand dan Vietnam. Data tahunan meliputi pendapatan bruto perkapita, arus masuk investasi asing, tingkat emisi karbon, jumlah konsumsi bahan bakar fosil dan total energi yang digunakan. Data ini diolah melalui proses prapemrosesan, pelatihan dengan model LSTM hingga pengujian sehingga menghasilkan model prediksi konsumsi energi baru terbarukan dengan arsitektur dan *hyperparameter* yang didapatkan dari hasil eksperimen. Hasil evaluasi model prediksi yaitu MAE dengan nilai 32,6503 dan R^2 sebesar 0,9556.

Kata kunci: Konsumsi energi, Energi Baru Terbarukan, Model prediksi, LSTM

IMPLEMENTATION OF LONG SHORT TERM MEMORY (LSTM) ALGORITHM FOR RENEWABLE ENERGY CONSUMPTION FORECASTING

ABSTRACT

Increased global energy consumption is driving countries around the world to look for alternative energy sources. With the limited availability of fossil fuels, renewable energy is the right solution to address global energy needs. However, not all countries have developed the technology to implement the use of renewable energy as an energy source. Therefore, in this study, an attempt was made to build a renewable energy consumption prediction model with deep learning. The data used is energy consumption data in developing countries in Asia including Bangladesh, Brunei Darussalam, China, Philippines, India, Indonesia, Iran, Cambodia, Laos, Malaysia, Myanmar, Pakistan, Thailand, and Vietnam. Vietnam. The annual data includes Gross Domestic Percapita (GDP), Foreign Direct Investment (FDI), carbon emission levels, total fossil fuel consumption and total energy used. This data is processed through the preprocessing, training process with the LSTM model and testing, resulting a renewable energy consumption prediction model with the architecture and hyperparameters obtained from the training process results. The prediction model evaluation results are MAE with a value of 32,6503 and R2 of 0,9556.

Keyword: Energy consumption, Renewable energy, Prediction model, LSTM

DAFTAR ISI

| | |
|--|------|
| PERSETUJUAN | iii |
| PERNYATAAN | iv |
| PENGHARGAAN | v |
| ABSTRAK | vii |
| <i>ABSTRACT</i> | viii |
| DAFTAR ISI | ix |
| DAFTAR GAMBAR | xii |
| DAFTAR TABEL | xiii |
| BAB 1 PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 3 |
| 1.3 Batasan Masalah | 3 |
| 1.4 Tujuan Penelitian | 4 |
| 1.5 Manfaat Penelitian | 4 |
| 1.6 Metodologi Penelitian | 4 |
| 1.7 Penelitian Relevan | 5 |
| 1.8 Sistematika Penulisan | 7 |
| BAB 2 TINJAUAN PUSTAKA | 8 |
| 2.1 Energi | 8 |
| 2.1.1 Energi Baru Terbarukan | 8 |
| 2.1.2 Analisis Konsumsi Energi | 9 |
| 2.2 Algoritma <i>Long Short Term Memory (LSTM)</i> | 11 |
| 2.2.1 Struktur Algoritma | 11 |
| 2.3 <i>Outliers</i> | 14 |
| 2.4 Imputasi Data Deret Waktu | 15 |
| 2.5 Normalisasi Data Deret Waktu | 15 |
| 2.6 Korelasi Variabel | 16 |
| 2.6.1 <i>Feature Importance</i> | 16 |
| 2.6.2 <i>Heatmap Correlation</i> | 16 |
| 2.7 <i>Hyperparameter</i> | 17 |
| 2.7.1 <i>Hidden Layer dan Nodes</i> | 17 |
| 2.7.2 <i>Epoch</i> | 17 |
| 2.7.3 <i>Fungsi Optimasi</i> | 17 |

| | |
|---|----|
| 2.7.4 <i>Learning rate</i> | 18 |
| 2.7.5 <i>Timestep</i> | 18 |
| 2.7.6 <i>Batch size</i> | 18 |
| 2.7.7 <i>Fungsi Aktivasi</i> | 18 |
| 2.7.8 <i>Dropout</i> | 19 |
| 2.8 <i>Group K-Fold Cross Validation</i> | 19 |
| 2.9 Metrik Evaluasi | 20 |
| 2.9.1 <i>Mean Absolute Error (MAE)</i> | 20 |
| 2.9.2 <i>R-squared (R^2)</i> | 20 |
| BAB 3 ANALISIS DAN PERANCANGAN SISTEM | 21 |
| 3.1 Analisis Sistem | 21 |
| 3.1.1 Analisis Masalah | 21 |
| 3.1.2 Analisis Kebutuhan | 22 |
| 3.2 Pemodelan Sistem | 23 |
| 3.2.1 Arsitektur Sistem | 23 |
| 3.2.2 <i>Use Case Diagram</i> | 24 |
| 3.2.3 <i>Sequence Diagram</i> | 24 |
| 3.2.4 <i>Activity Diagram</i> | 25 |
| 3.2.5 <i>Flowchart</i> | 26 |
| 3.3 Analisis Proses | 27 |
| 3.3.1 Pengumpulan Data | 27 |
| 3.3.2 Perhitungan Prediksi Konsumsi Energi Baru Terbarukan dengan Algoritma LSTM | 29 |
| 3.4 <i>Pseudocode</i> | 38 |
| 3.5 Perancangan Antarmuka Pengguna | 39 |
| 3.5.1 Halaman awal | 39 |
| 3.5.2 Halaman input | 39 |
| 3.5.3 Halaman output | 40 |
| 3.5.4 Halaman akhir | 41 |
| BAB 4 HASIL DAN PEMBAHASAN | 42 |
| 4.1 Spesifikasi Perangkat Keras | 42 |
| 4.2 Hasil Prapemrosesan Data | 42 |
| 4.2.1 Hasil Penanganan <i>Outliers</i> | 42 |
| 4.2.2 Hasil Penanganan <i>Missing Value</i> | 43 |

| | |
|---|----|
| 4.2.3 Hasil Normalisasi | 44 |
| 4.2.4 Pengecekan Korelasi Variabel | 45 |
| 4.3 Analisis Arsitektur Model | 47 |
| 4.3.1 Eksplorasi <i>hyperparameter</i> | 47 |
| 4.3.2 Arsitektur Model | 53 |
| 4.3 Hasil Validasi Menggunakan <i>Group K-Fold Cross Validation</i> | 54 |
| 4.4 Hasil Evaluasi Model | 55 |
| 4.4.1 <i>Mean Absolute Error (MAE)</i> | 55 |
| 4.4.2 <i>R-squared (R²)</i> | 55 |
| 4.5 Hasil Prediksi Konsumsi Energi Baru Terbarukan..... | 55 |
| 4.6 Hasil Pengujian Model Prediksi..... | 56 |
| 4.7 Hasil Integrasi Sistem..... | 57 |
| 4.7.1 Halaman awal..... | 57 |
| 4.7.2 Halaman input | 58 |
| 4.7.3 Halaman output | 60 |
| 4.7.4 Halaman akhir | 61 |
| BAB 5 PENUTUP | 62 |
| 5.1 Kesimpulan..... | 62 |
| 5.2 Saran..... | 62 |
| DAFTAR PUSTAKA..... | 63 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 1 Tingkat konsumsi bahan bakar fosil di dunia..... | 9 |
| Gambar 2 Tingkat konsumsi bahan bakar fosil di Indonesia | 10 |
| Gambar 3 Produksi pembangkit listrik terbarukan di Indonesia..... | 10 |
| Gambar 4 Struktur algoritma LSTM..... | 11 |
| Gambar 5 <i>Fishbone</i> diagram..... | 21 |
| Gambar 6 Arsitektur sistem | 23 |
| Gambar 7 <i>Use case</i> diagram | 24 |
| Gambar 8 <i>Sequence</i> diagram | 24 |
| Gambar 9 <i>Activity</i> diagram | 25 |
| Gambar 10 <i>Flowchart</i> sistem..... | 26 |
| Gambar 11 <i>Flowchart</i> algoritma LSTM | 27 |
| Gambar 12 Halaman awal..... | 39 |
| Gambar 13 Halaman input | 39 |
| Gambar 14 Halaman output | 40 |
| Gambar 15 Halaman output grafik | 40 |
| Gambar 16 Halaman akhir | 41 |
| Gambar 17 Tingkat pengaruh fitur terhadap target..... | 45 |
| Gambar 18 <i>Heatmap correlation</i> | 46 |
| Gambar 19 Grafik <i>training loss</i> dan <i>validation loss</i> | 49 |
| Gambar 20 Halaman awal web | 57 |
| Gambar 21 Panduan pengisian data | 58 |
| Gambar 22 Pesan kesalahan pada halaman input | 58 |
| Gambar 23 Halaman input web | 59 |
| Gambar 24 Pesan prediksi berhasil..... | 59 |
| Gambar 25 Halaman output web | 60 |
| Gambar 26 Grafik tren konsumsi energi baru terbarukan..... | 60 |
| Gambar 27 Halaman akhir web | 61 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 1 Rincian dataset | 28 |
| Tabel 2 Sampel data | 29 |
| Tabel 3 Transformasi log..... | 30 |
| Tabel 4 Penanganan missing value dengan <i>backward fill</i> | 33 |
| Tabel 5 Normalisasi data..... | 34 |
| Tabel 6 Nilai input, bobot dan bias | 34 |
| Tabel 7 Spesifikasi perangkat keras | 42 |
| Tabel 8 Jumlah <i>outliers</i> yang ditangani | 43 |
| Tabel 9 Jumlah <i>missing value</i> yang ditangani..... | 43 |
| Tabel 10 Hasil prapemrosesan data..... | 44 |
| Tabel 11 Hasil eksplorasi jumlah <i>hidden layer</i> dan <i>nodes</i> | 48 |
| Tabel 12 Hasil eksplorasi nilai parameter <i>patience</i> | 49 |
| Tabel 13 Hasil eksplorasi fungsi optimasi | 50 |
| Tabel 14 Hasil eksplorasi nilai <i>timestep</i> | 51 |
| Tabel 15 Hasil eksplorasi nilai <i>batch size</i> | 51 |
| Tabel 16 Hasil eksplorasi fungsi aktivasi..... | 52 |
| Tabel 17 Hasil eksplorasi nilai <i>dropout</i> | 52 |
| Tabel 18 <i>Model summary</i> | 53 |
| Tabel 19 <i>Hyperparameter</i> | 53 |
| Tabel 20 Hasil <i>group k-fold cross validation</i> | 54 |
| Tabel 21 Hasil prediksi konsumsi energi baru terbarukan | 55 |
| Tabel 22 Tabel perbandingan nilai aktual dan nilai prediksi..... | 56 |
| Tabel 23 Hasil pengujian model prediksi..... | 57 |

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Sebagian besar energi yang digunakan dunia saat ini merupakan energi yang dihasilkan dari pengolahan bahan bakar fosil seperti batu bara, minyak bumi dan gas alam. Menurut data yang dirilis oleh *Our World in Data*, awalnya hanya batu bara yang marak digunakan. Lalu setelah beberapa puluh tahun, sumber energi beralih dengan penggunaan minyak dan gas alam. Konsumsi bahan bakar fosil di seluruh dunia terus meningkat hingga mencapai 140,231 TWh per tahun 2023 (Ritchie & Rosado, 2024).

Dalam penelitian yang telah dilakukan oleh Meng et al., di tahun 2022, prediksi konsumsi energi di China akan melonjak hingga 7% yaitu sebesar 6 miliar ton batu bara pada tahun 2024 (Meng et al., 2022). Penelitian lainnya menyebutkan bahwa penggunaan bahan bakar fosil sebagai sumber energi juga masih dominan di Indonesia, terutama di pulau Sumatera, Jawa dan Kalimantan (Pambudi et al., 2023). Menurut *Handbook of Energy and Economic Statistics of Indonesia* Tahun 2023 oleh Kementerian Energi dan Sumber Daya Mineral Republik Indonesia, terjadi peningkatan konsumsi energi hingga 6,29% dimana industri menjadi sektor dengan permintaan energi terbesar (Kementerian Energi dan Sumber Daya Mineral, 2023).

Menurut artikel yang dipublikasi oleh *HowStuffWorks*, apabila konsumsi bahan bakar fosil di seluruh dunia terus meningkat mungkin cadangan minyak masih ada namun harganya akan terus melonjak karena banyaknya permintaan. Pada akhirnya, orang-orang termasuk pemerintah akan mencari cara alternatif dengan menggunakan energi baru terbarukan yang lebih terjangkau dan ramah lingkungan (Flinn, 2024). Namun, proses adaptasi dan transisi dari penggunaan bahan bakar fosil menuju energi baru terbarukan tidak dapat dilakukan dalam waktu yang sebentar. Penelitian terdahulu menyatakan bahwa transisi akan terasa sangat lambat khususnya di Indonesia karena masih sangat sedikit investasi penelitian dan pengembangan teknologi energi baru terbarukan yang tersedia di dalam negeri (Udin, 2020). Selain itu, pola konsumsi energi baru terbarukan cenderung belum dapat

dipastikan karena akan dipengaruhi oleh banyak faktor seperti konsumsi listrik, konsumsi bahan bakar fosil, arus masuk investasi hingga pertumbuhan ekonomi (Hoa et al., 2023).

Dengan perkembangan data dan algoritma saat ini, teknologi dapat dimanfaatkan sebagai alat untuk melakukan analisis untuk tujuan klasifikasi, deteksi hingga prediksi masa depan dalam berbagai bidang termasuk energi. Pedro Chévez dan Irene Martini bereksperimen pada penelitiannya dengan menggunakan ANN untuk memprediksi konsumsi listrik pada bangunan universitas. Hasil penelitian menunjukkan bahwa ANN bekerja lebih baik pada prediksi jangka pendek (per jam) dibandingkan dengan prediksi jangka panjang (per tahun) (Chévez & Martini, 2024). Penelitian lainnya menyimpulkan bahwa dari pengujian dengan menggunakan tiga dataset historis yang berbeda ditemukan bahwa algoritma LightGBM, XGBoost dan LSTM menjadi algoritma paling mutakhir untuk membangun model prediksi (Dolgintseva et al., 2024). Talwariya et al., juga menyebutkan bahwa LSTM menjadi algoritma yang terbukti efektif untuk melakukan prediksi konsumsi energi dengan data historis (Talwariya et al., 2023). LSTM merupakan implementasi dari *Recurrent Neural Network (RNN)* yang pertama kali diusulkan oleh Hochreiter dan Schmidhuber pada tahun 1997. LSTM mampu mempelajari pola jangka panjang yang kompleks dengan waktu yang lebih singkat dibandingkan dengan algoritma RNN yang telah ada sebelumnya (Hochreiter & Schmidhuber, 1997). Model LSTM sengaja dirancang untuk menangani data deret waktu dengan kemampuannya untuk mengatasi masalah gradien yang hilang (Van Houdt et al., 2020).

Penelitian ini menggunakan LSTM dalam memprediksi tren yang terjadi dalam konsumsi listrik yang bersumber dari energi baru terbarukan. Implementasi dari penelitian ini diharapkan dapat memberikan wawasan yang berguna bagi pemangku kepentingan, seperti pemerintah dalam merumuskan kebijakan energi yang tepat serta perusahaan energi dalam menyusun strategi bisnis dan perencanaan yang lebih efektif. Sebagai contoh, hasil penelitian ini dapat membantu pemerintah merancang dan membangun infrastruktur untuk meningkatkan angka konsumsi energi terbarukan, sementara perusahaan penyedia panel surya dapat memfokuskan layanannya pada area dengan

permintaan yang tinggi sesuai dengan prediksi yang dihasilkan. Selain itu, analisis yang dihasilkan juga diharapkan dapat menjadi acuan bagi penelitian di masa depan dalam mempertimbangkan penggunaan algoritma LSTM.

1.2 Rumusan Masalah

Dengan tingginya permintaan energi dan terbatasnya sumber energi dari bahan bakar fosil, energi baru terbarukan menjadi solusi praktis untuk memecahkan masalah tersebut. Namun terdapat masalah dalam proses peralihan bahan bakar fosil menuju energi baru terbarukan yaitu pola konsumsi energi baru terbarukan yang tidak pasti. Oleh karena itu, diperlukan suatu model prediksi yang akurat dan efektif untuk memperkirakan konsumsi energi baru terbarukan di masa depan.

1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut

1. Prediksi yang dilakukan dalam penelitian ini adalah prediksi konsumsi listrik dari energi baru terbarukan per orang per tahun pada 14 negara berkembang di Asia yang bersumber dari *World Bank Open Data* dan *Our World in Data*.
2. Parameter yang digunakan adalah data tahunan dari 1974 hingga 2023 yang mencakup data Pendapatan Bruto Perkapita atau *Gross Domestic Product (GDP)* dalam USD, arus masuk investasi asing atau *Foreign Direct Investment (FDI)* dalam USD, tingkat emisi karbon (ton/orang), jumlah konsumsi energi fosil (kWh/orang), total energi yang digunakan (kWh/orang) dan jumlah konsumsi energi baru terbarukan (kWh/orang).
3. Model dibangun dengan bahasa pemrograman *Python* dan akan diimplementasikan dalam web dengan menggunakan *streamlit*.
4. Algoritma yang digunakan untuk melakukan prediksi konsumsi energi baru terbarukan adalah algoritma *Long Short Term Memory (LSTM)*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk membangun sistem prediksi konsumsi energi baru terbarukan guna mendukung pengambilan keputusan untuk masa yang akan datang.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah untuk dapat memberikan informasi prediksi konsumsi energi baru terbarukan untuk membantu pemangku kepentingan dalam mengambil keputusan dan membuat perencanaan masa depan. Juga dapat memberikan informasi terkait kinerja algoritma LSTM dalam sistem prediksi dengan data tahunan konsumsi energi baru terbarukan sehingga dapat menjadi referensi bagi penelitian berikutnya

1.6 Metodologi Penelitian

Adapun metodologi yang digunakan dalam penelitian ini adalah sebagai berikut

1. Studi pustaka

Penelitian dimulai dengan mencari referensi dari berbagai sumber terpercaya dan melakukan studi pustaka melalui buku-buku, jurnal, *e-book*, artikel ilmiah, makalah ataupun situs internet yang berhubungan dengan energi baru terbarukan dan algoritma *Long Short Term Memory (LSTM)*

2. Analisis dan perancangan

Pada tahap analisis, penulis menganalisis terkait fenomena energi baru terbarukan yang sedang marak digaungkan demi menghindari habisnya bahan bakar fosil. Penulis juga mengidentifikasi apa saja yang akan dibutuhkan dalam penelitian untuk segera dirancang dalam sebuah diagram alir (*flowchart*).

3. Pengumpulan dan prapemrosesan data

Data yang digunakan dalam penelitian ini adalah data konsumsi energi tahunan negara-negara berkembang di Asia. Data diekstrak sesuai dengan kebutuhan penelitian dengan menyertakan sitasi kepada penyedia data. Kemudian dilakukan tahap prapemrosesan data seperti penanganan nilai kosong dan pembersihan data.

4. Implementasi

Proses pembuatan model diawali dengan melakukan normalisasi data input untuk memastikan setiap fitur memiliki skala yang sama. Model dilatih dengan beberapa *epoch* untuk mencapai performa terbaik berdasarkan hasil validasi pada data uji.

5. Pengujian dan integrasi sistem

Pada tahap pengujian model, dilakukan evaluasi terhadap performa model LSTM yang telah dilatih menggunakan data uji yang terpisah dari data pelatihan. Evaluasi dan analisis harus dilakukan untuk memperoleh pemahaman terkait kinerja algoritma dalam melakukan tugasnya. Setelah mendapatkan kinerja model yang diharapkan, penulis kemudian mengintegrasikan model tersebut ke dalam tampilan web yang nantinya dapat diakses oleh pengguna.

6. Dokumentasi

Pada tahap ini, dilakukan dokumentasi penelitian mulai dari tahap analisa sampai kepada analisis untuk kemudian disusun dalam bentuk skripsi.

1.7 Penelitian Relevan

1. Hasil analisis dalam penelitian berjudul “*Forecasting energy consumption based on SVR and Markov model: A case study of China*” yang telah dilakukan oleh Meng, Z., Sun, H., & Wang, X. pada tahun 2022 menyatakan bahwa konsumsi energi di Cina akan terus meningkat seiring dengan industrialisasi dan urbanisasi. Peneliti memanfaatkan SVR dan *Markov Chain* untuk memprediksi struktur konsumsi energi dengan akurasi lebih dari 98,4%. Sebagai hasilnya, peneliti mendapatkan kesimpulan bahwa konsumsi energi di Cina akan meningkat rata-rata 7% per tahun dengan jumlah penggunaan batu bara melebihi 6 miliar ton pada tahun 2024.
2. Dalam penelitian berjudul “*Machine learning based renewable energy generation and energy consumption forecasting*” yang disusun oleh Talwariya, A., Singh, P., Jobanputra, J. H., & Kolhe, M. L. pada tahun 2023 disimpulkan bahwa algoritma LSTM bekerja dengan baik untuk data historis yang bersifat sekuensial. Peneliti menggunakan algoritma LSTM untuk melakukan prediksi konsumsi energi dan ini memberikan hasil evaluasi yang

baik daripada algoritma CNN yang digunakan untuk memprediksi energi cahaya matahari dan angin. Penulis merekomendasikan untuk meningkatkan akurasi model dengan memodifikasi nilai *hyperparameter* dan menggunakan dataset yang mencakup lebih banyak jenis data

3. Hasil eksperimen yang dilakukan oleh Gehad Ismail Sayed, Eman I. Abd El-Latif, Aboul Ella Hassanien dan Vaclav Snasel dari penelitian yang berjudul *“Optimized long short-term memory with rough set for sustainable forecasting renewable energy generation”* menemukan model pengembangan LSTM bernama RSNOA (*Rough Set and Nutracker Optimization Algorithm*) yang mampu memprediksi produksi energi baru terbarukan. Model ini memiliki nilai MAE sebesar 2.835 yang membuktikan keakuratannya dalam melakukan prediksi.
4. Berdasarkan temuan yang didapatkan oleh Hoa, P. X., Xuan, V. N., & Thu, N. T. P. dalam penelitiannya yang berjudul *“Determinants of the renewable energy consumption: The case of Asian countries”* pada tahun 2023 disimpulkan bahwa ada beberapa faktor yang memengaruhi konsumsi energi baru terbarukan di negara-negara Asia yaitu pertumbuhan ekonomi, konsumsi listrik, penggunaan bahan bakar fosil dan arus investasi asing.
5. Berdasarkan penelitian berjudul *“Renewable Energy in Indonesia: Current Status, Potential, and Future Development”* yang disusun oleh Pambudi, N. A., Firdaus, R. A., Rizkiana, R., Ulfa, D. K., Salsabila, M. S., Suharno, & Sukatiman pada tahun 2023, ditemukan bahwa Indonesia masih bergantung pada bahan bakar fosil seperti batu bara, minyak dan gas alam. Penulis juga menyatakan bahwa Indonesia memiliki potensi yang tinggi untuk pengembangan energi baru terbarukan yaitu sebesar 419 GW. Penulis memberi saran untuk fokus pada masa transisi ke energi baru terbarukan dengan meningkatkan efisiensi energi seperti mendorong penggunaan kendaraan listrik dan pembangkit listrik tenaga surya, juga mengembangkan teknologi yang mendorong kemajuan energi baru terbarukan.
6. Pada penelitian berjudul *“Near real-time wind speed forecast model with bidirectional LSTM networks”* yang dilakukan oleh Lionel P. Joseph, Ravinesh C. Deo, Ramendra Prasad, Sancho Salcedo-Sanz, Nawin Raj dan Jeffrey Soar ditemukan bahwa model Bi-LSTM yang diusulkan

menunjukkan kinerja superior dibandingkan LSTM standar, MLP, RNN dan RF. Bi-LSTM mencapai hasil terbaik dalam pengujian dengan MAPE antara 8,8% hingga 21,5%. Penelitian ini menemukan bahwa cuaca menjadi faktor penting dalam memprediksi kecepatan angin.

1.8 Sistematika Penulisan

Adapun sistematika penulisan skripsi ini terdiri dari beberapa bagian, yaitu:

BAB 1 PENDAHULUAN

Bab ini berisi tentang latar belakang, rumusan masalah, batasan penelitian, tujuan dan manfaat penelitian, penelitian relevan, metodologi penelitian dan sistematika penulisan

BAB 2 TINJAUAN PUSTAKA

Pada bab ini dijelaskan tentang energi, algoritma *Long Short Term Memory (LSTM)*, *hyperparameter*, *k-fold cross validation* dan juga metrik evaluasi yang digunakan untuk menguji keakuratan model prediksi.

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Bab ini berisi tentang analisis kebutuhan, arsitektur sistem, pemodelan sistem, analisis proses, *pseudocode* dan perancangan antarmuka pengguna.

BAB 4 HASIL DAN PEMBAHASAN

Bab ini akan menjelaskan tentang hasil penelitian hingga tahap pengujian.

BAB 5 KESIMPULAN DAN SARAN

Bab ini mencakup kesimpulan yang didapatkan dari penelitian dan saran dari peneliti yang bisa membantu penelitian selanjutnya.

BAB 2

TINJAUAN PUSTAKA

2.1 Energi

Sumber daya menjadi salah satu bagian penting untuk keberlangsungan hidup. Energi yang dihasilkan dari sumber daya ini dapat digunakan untuk berbagai aktivitas seperti menerangi ruangan dan menggerakkan mesin. Indonesia kaya akan sumber energi potensial, baik itu berupa minyak dan batu bara yang termasuk ke dalam jenis bahan bakar fosil maupun sumber energi baru terbarukan seperti cahaya matahari, angin, air, gelombang laut, biomassa dan panas bumi (Fikriyyah & Boedoyo, 2021).

Namun hingga saat ini sebagian besar negara di dunia masih bergantung pada bahan bakar fosil untuk menghasilkan listrik (Ang et al., 2022). Penggunaan energi fosil memberi dampak negatif yang mendorong perubahan iklim global. Batu bara, minyak, gas dan semua jenis bahan bakar fosil akan mengeluarkan karbondioksida ketika dibakar. Peningkatan penggunaan energi yang tidak ramah lingkungan juga berkontribusi pada peningkatan emisi karbon (Nonthapot et al., 2024). Selain menjadi faktor penyebab terjadinya perubahan iklim, penggunaan bahan bakar fosil sebagai sumber energi juga merupakan penyebab tingginya polusi udara hingga menjadi alasan kematian dini setiap tahunnya (Ritchie & Rosado, 2024).

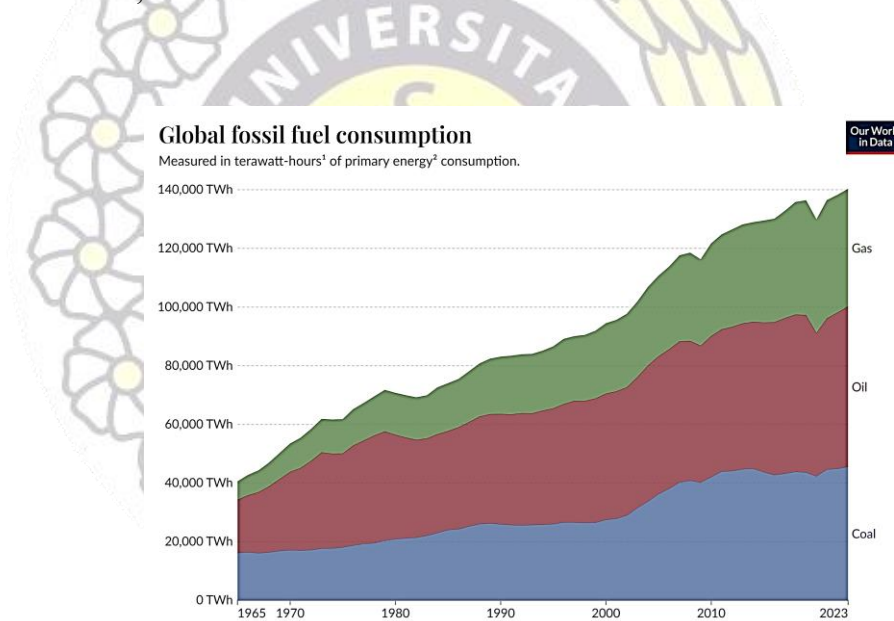
2.1.1 Energi Baru Terbarukan

Energi baru terbarukan meliputi beberapa sumber daya yaitu tenaga surya, angin, air, biomassa dan panas bumi. Lokasi yang strategis dan iklim tropis dengan matahari sepanjang tahun menjadikan Asia Tenggara sebagai wilayah dengan potensi yang besar untuk pengembangan sumber daya panas bumi dan tenaga surya (Pratiwi & Juerges, 2020). Namun, meskipun Indonesia memiliki sumber energi baru terbarukan yang potensial, butuh waktu dan perencanaan yang lama untuk dapat memanfaatkan energi baru terbarukan secara optimal karena biaya investasi teknologi yang tinggi (Udin, 2020). Tetapi, adanya Rencana Umum Ketenagalistrikan Nasional 2023-2060 membuktikan bahwa Indonesia sebagai negara berkembang

terus menyusun dan melakukan perencanaan untuk memberdayakan energi baru terbarukan demi memenuhi kebutuhan listrik nasional (Minister of Energy and Mineral Resources of the Republic of Indonesia, 2023). Hingga pertengahan tahun 2024, bauran energi baru terbarukan di Indonesia mencapai 13,93% dari target sebesar 23% pada tahun 2025 (Adi, 2024).

2.1.2 Analisis Konsumsi Energi

Berdasarkan Gambar 1, dapat diketahui bahwa konsumsi sumber daya tidak terbarukan di dunia terus meningkat secara signifikan setiap tahunnya. Penggunaan batu bara dan minyak lebih banyak dibandingkan dengan penggunaan gas alam. Hingga tahun 2023, penggunaan energi fosil menyentuh angka konsumsi sebanyak 140,000 TWh di seluruh dunia.

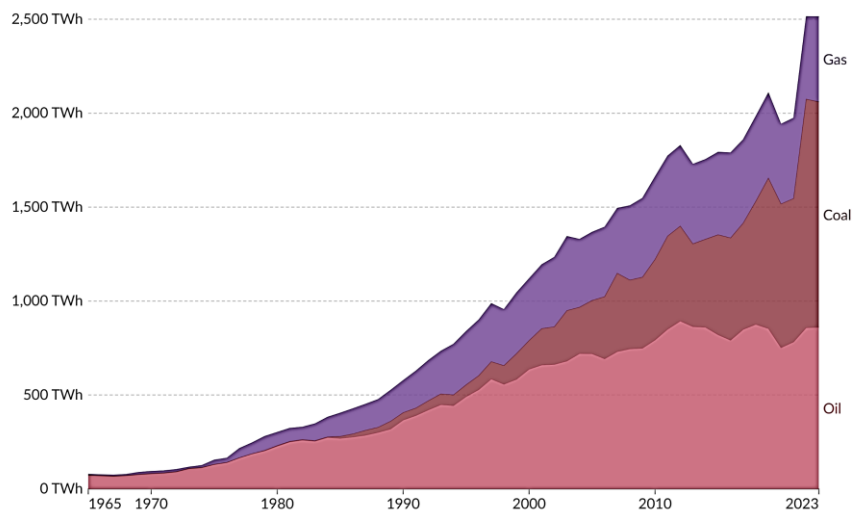


Gambar 1 Tingkat konsumsi bahan bakar fosil di dunia

Sama halnya dengan Gambar 2 yang menunjukkan bahwa minyak bumi dan batu bara juga menjadi dua tipe bahan bakar fosil yang paling banyak digunakan di Indonesia dengan jumlah konsumsi lebih dari 2,500 TWh. Tren kedua grafik ini menggambarkan bahwa konsumsi bahan bakar fosil terus mengalami peningkatan yang signifikan hingga tahun 2023.

Fossil fuel consumption by fuel type, Indonesia

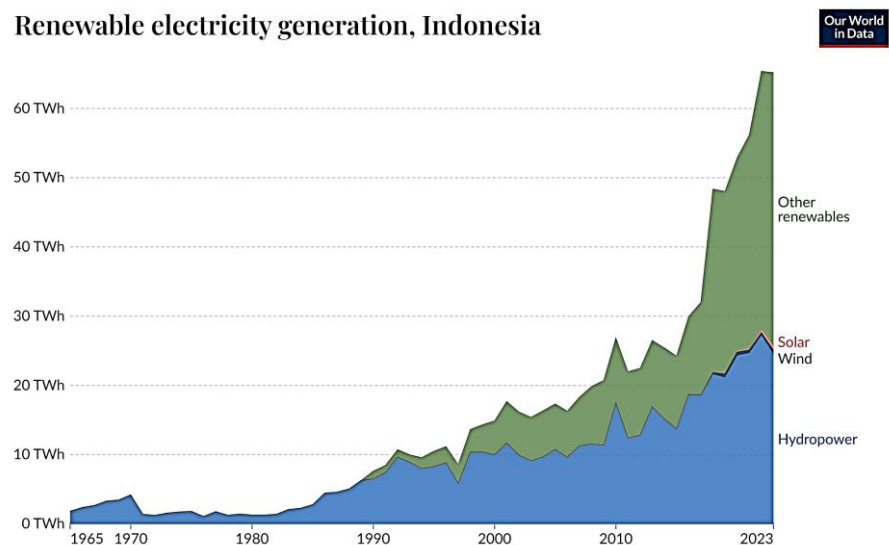
Fossil fuel consumption is given in terawatt-hours (TWh).



Gambar 2 Tingkat konsumsi bahan bakar fosil di Indonesia

Salah satu hal yang menjadi penyebab tren konsumsi energi fosil terus meningkat adalah subsidi energi fosil. Ketika harga bahan bakar fosil meningkat, pemerintah mengambil langkah untuk meningkatkan jumlah subsidi untuk mengurangi beban masyarakat. Namun langkah ini secara tidak langsung menyebabkan ketergantungan jangka panjang dalam menggunakan energi fosil (Muta & Erdogan, 2023).

Renewable electricity generation, Indonesia



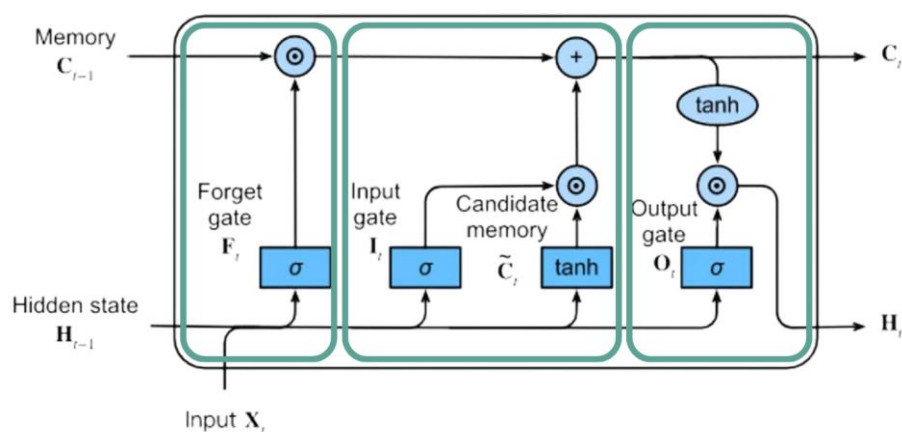
Gambar 3 Produksi pembangkit listrik terbarukan di Indonesia

Sebagai perbandingan, Gambar 3 merupakan grafik produksi listrik dari pembangkit listrik terbarukan di Indonesia. Jumlah produksi tertinggi hanya mencapai kurang dari 70 TWh, sedangkan mengacu pada Gambar 2 tertera bahwa konsumsi bahan bakar fosil mencapai lebih dari 2,500 TWh. Hal ini menyiratkan bahwa masih perlu persiapan dan pemberdayaan energi baru terbarukan hingga dapat memenuhi kebutuhan energi nasional. Pambudi et al., menyampaikan bahwa dengan data konsumsi energi saat ini, penggunaan bahan bakar fosil masih dominan dan akan terus meningkat (Pambudi et al., 2023). Maka diperlukan strategi, investasi dan perencanaan yang matang untuk pemanfaatan energi terbarukan yang tak terbatas secara optimal di masa yang akan datang.

2.2 Algoritma *Long Short Term Memory (LSTM)*

LSTM merupakan algoritma lanjutan dari RNN yang dibuat untuk mengatasi isu *vanishing gradient* yang sering terjadi pada algoritma RNN sebelumnya seperti *Backpropagation Through Time (BPTT)* atau *Real Time Recurrent Learning (RTRL)*. LSTM bekerja dengan efektif untuk data yang memerlukan memori jangka panjang seperti prediksi deret waktu, pengenalan suara dan pemrosesan bahasa alami. Arsitekturnya dapat dengan mudah dimodifikasi untuk tugas yang berbeda-beda tanpa kehilangan kemampuannya untuk mengelola dan menyimpan informasi jangka panjang (Hochreiter & Schmidhuber, 1997).

2.2.1 Struktur Algoritma



Gambar 4 Struktur algoritma LSTM

Adapun rincian tentang struktur LSTM (Hochreiter & Schmidhuber, 1997) disederhanakan sebagai berikut:

1. **Forget gate (f_t):** berfungsi untuk mengatur seberapa banyak informasi dari langkah sebelumnya yang harus dilupakan.

$$f_t = \sigma(W_f \cdot X_t + W_f \cdot h_{t-1} + b_f) \dots \dots \dots (1)$$

Hasil kali antara bobot (W_f) dan input (X_t) kemudian dijumlahkan dengan hasil kali bobot (W_f) dan *hidden state* sebelumnya (h_{t-1}) dan dijumlahkan dengan bias (b_f) lalu hasilnya akan melewati fungsi aktivasi sigmoid (σ). Fungsi ini akan menghasilkan nilai antara 0 dan 1 dimana nilai 0 berarti informasi dilupakan sepenuhnya dan nilai 1 berarti informasi dipertahankan atau “diingat” sepenuhnya.

2. **Input gate (i_t):** berfungsi untuk menentukan informasi baru yang akan dimasukkan ke dalam *cell state* pada waktu t .

$$i_t = \sigma(W_i X_t + W_i h_{t-1} + b_i) \dots \dots \dots (2)$$

Tahap ini bertujuan untuk mengontrol informasi baru yang akan diikutsertakan dalam struktur berikutnya. Sama seperti *forget gate* (f_t), hasil operasi antara bobot (W_i), input (X_t), *hidden state* sebelumnya (h_{t-1}) dan bias (b_i) akan melewati fungsi sigmoid (σ). Output dari fungsi sigmoid ini kemudian akan mengatur seberapa banyak informasi yang akan disimpan untuk tahap berikutnya.

3. **Candidate cell state (\hat{C}_t):** berfungsi untuk menyediakan dan menyiapkan informasi baru dari hasil pada tahap *input gate*

$$\hat{C}_t = \tanh(W_c X_t + W_c h_{t-1} + b_c) \dots \dots \dots (3)$$

Tahap ini dihasilkan oleh operasi bobot (W_c), input (X_t), *hidden state* sebelumnya (h_{t-1}) dan bias (b_c) yang melewati fungsi aktivasi tanh. Output dari fungsi tersebut akan menghasilkan nilai diantara 1 dan -1. Output mendekati 1 berarti menambah informasi, output -1 berarti mengurangi atau mengubah informasi dan output 0 berarti tidak menambah maupun mengurangi informasi yang ada.

4. **Cell state update (c_t):** berfungsi untuk menyimpan nilai yang telah diubah pada *gate* sebelumnya

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \hat{C}_t \dots \dots \dots (4)$$

Cell state (c_t) adalah bagian paling penting dari LSTM, karena berfungsi sebagai tempat informasi disimpan dari waktu ke waktu. Persamaan ini melakukan dua hal:

- $f_t c_{t-1}$: Mengatur bagian mana dari *cell state* sebelumnya c_{t-1} yang akan dilupakan. Hasil ini berasal dari *forget gate* yang dikalikan dengan nilai *cell state* sebelumnya.
- $i_t \hat{C}_t$: Menentukan seberapa banyak informasi baru \hat{C}_t yang akan ditambahkan ke *cell state*, tergantung pada output dari *input gate*.

5. **Output gate (O_t):** berfungsi untuk menentukan informasi apa yang akan diambil dari *cell state* sebagai output

$$O_t = \sigma(W_o \cdot X_t + W_o \cdot h_{t-1} + b_o) \dots \dots \dots (5)$$

Output gate menentukan bagian dari *cell state* yang akan digunakan sebagai output pada langkah waktu saat ini. Output dari *output gate* akan menentukan seberapa banyak dari *cell state* yang akan diteruskan sebagai *hidden state*.

6. **Hidden state (H_t):** berfungsi untuk menghasilkan output dari seluruh proses yang telah dilakukan oleh LSTM

$$H_t = O_t \cdot \tanh(C_t) \dots \dots \dots (6)$$

Hidden state merupakan tahap dimana LSTM akan menghasilkan output akhir. Tahap ini dihasilkan dari operasi antara *output gate* dan *cell state* yang telah diaktivasi oleh fungsi tanh. Ini mengatur bagian mana dari *cell state* yang akan dikirim ke luar.

2.3 Outliers

Pencilan (*outliers*) adalah data yang memiliki nilai yang sangat berbeda dibandingkan dengan sebagian besar data lainnya. Keberadaan *outlier* bisa disebabkan oleh kesalahan dalam pengukuran, pemrosesan data, atau variasi data dalam dataset. Kehadiran *outlier* dapat memengaruhi hasil analisis statistik dan performa model sehingga penting untuk mengidentifikasi dan menangani data tersebut secara efektif. Salah satu metode yang sering digunakan untuk mendeteksi *outlier* adalah *Interquartile Range (IQR)* (Dash et al., 2023). Deteksi *outlier* dengan IQR dapat digambarkan secara matematis sebagai berikut:

1. Hitung posisi kuartil atas (Q1) dan kuartil bawah (Q3). Untuk menentukan Q1 dan Q3, susun data dari nilai terkecil hingga nilai terbesar (*ascending order*) dan hitung jumlah data (n). Q1 dan Q3 dapat ditentukan berdasarkan persamaan berikut:

$$Q1 = 0,25 \times (n + 1) \dots\dots\dots(7)$$

$$Q3 = 0,75 \times (n + 1) \dots\dots\dots(8)$$

2. Hitung *Interquartile Range (IQR)*, yaitu selisih antara Q3 dan Q1.

$$IQR = Q3 - Q1 \dots\dots\dots(9)$$

3. Hitung batas bawah (*lower bound*) dan batas atas (*upper bound*) yang menjadi batas rentang data. Persamaan untuk menentukan batas rentang data dapat dijelaskan sebagai berikut:

$$Lower\ bound = Q1 - 1.5 \times IQR \dots\dots\dots(10)$$

$$Upper\ bound = Q3 + 1.5 \times IQR \dots\dots\dots(11)$$

Setelah mengetahui batas rentang data, dapat ditentukan bahwa data yang berada di luar batas bawah dan atas dapat dikategorikan sebagai *outlier*. Untuk mengatasi *outlier* yang ditemukan, metode *Winsorizing* dapat digunakan dengan mengganti nilai *outlier* dengan batas rentang data yang telah ditentukan (Obatola & Junjie, 2024). Dalam metode ini, nilai yang terdeteksi lebih kecil dari batas bawah (*lower bound*) atau lebih besar dari batas atas (*upper bound*) akan diganti dengan batas itu sendiri. Teknik *Winsorizing* membantu menormalkan dataset tanpa harus menghapus data apa pun, sehingga struktur dataset tetap terjaga.

2.4 Imputasi Data Deret Waktu

Dalam analisis deret waktu, *missing value* dapat terjadi karena keterbatasan data yang tersedia. Adanya *missing value* dapat mengganggu analisis dan pemodelan sehingga perlu ditangani dengan tepat. Dikutip dari laman Medium yang berjudul “*Missing Data Imputation Approaches and Application*”, metode imputasi data yang umum digunakan untuk data deret waktu antara lain:

1. Interpolasi linier: memperkirakan nilai yang hilang menggunakan garis lurus yang menghubungkan dua titik data terdekat. Metode ini sederhana, namun kurang akurat untuk data non-linear.
2. *Forward fill*: mengganti nilai yang hilang dengan nilai data sebelumnya. Metode ini efektif untuk mengisi nilai yang hilang yang tidak mempengaruhi tren jangka panjang.
3. *Backward fill*: mengganti nilai yang hilang dengan nilai data setelahnya (Kocatas, 2024).

2.5 Normalisasi Data Deret Waktu

Penskalaan deret waktu memiliki dampak signifikan terhadap kinerja model. Rentang nilai yang besar dapat menyebabkan proses pembelajaran berjalan lambat atau bahkan gagal karena ketidakstabilan. Selain itu, pada kasus dengan banyak input yang memiliki skala berbeda, input dengan rentang data yang lebih besar akan mendominasi input lainnya, terutama pada metode yang sensitif terhadap skala data. *MinMax scaling* adalah salah satu teknik normalisasi yang umum digunakan untuk mengatasi masalah ini (Meisenbacher et al., 2022). Teknik *MinMax scaling* bertujuan untuk mentransformasi data deret waktu ke dalam rentang nilai antara 0 hingga 1. Secara matematis dapat diuraikan sebagai berikut:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \dots \dots \dots (12)$$

Keterangan:

X_{scaled} : nilai setelah normalisasi

X : nilai asli data

X_{min} : nilai minimum dalam kolom

X_{max} : nilai maksimum dalam kolom

Proses ini dilakukan dengan mengurangi setiap nilai data dengan nilai minimum dari seluruh data, kemudian membaginya dengan selisih antara nilai maksimum dan nilai minimum. Dengan demikian, semua deret waktu akan memiliki rentang nilai yang sama, yang akan memudahkan proses perbandingan dan analisis.

2.6 Korelasi Variabel

2.6.1 *Feature Importance*

Feature importance ranking merupakan teknik yang krusial dalam pengembangan model jaringan saraf. Teknik ini bertujuan untuk mengidentifikasi kontribusi relatif dari setiap fitur dalam proses pengambilan keputusan model (Heidari et al., 2023). *Feature importance ranking* menggunakan *random forest* untuk melakukan pemeringkatan pentingnya fitur. *Random forest* yang terdiri dari sejumlah pohon keputusan, digunakan untuk menghitung nilai pentingnya setiap node berdasarkan seberapa baik node tersebut dapat memisahkan data ke dalam kelas yang berbeda. Dalam konteks jaringan saraf, *feature importance ranking* tidak hanya berfungsi untuk meningkatkan akurasi, tetapi juga untuk menyederhanakan model dengan mengeliminasi fitur yang kurang penting.

2.6.2 *Heatmap Correlation*

Heatmap correlation merupakan metode visualisasi yang digunakan untuk menunjukkan hubungan antara variabel dalam dataset dengan menggunakan warna untuk merepresentasikan tingkat korelasi antar variabel. Dalam konteks pemodelan prediktif, *heatmap correlation* membantu dalam mengidentifikasi variabel input yang paling relevan dan signifikan terhadap variabel output yang ingin diprediksi. Proses ini dimulai dengan menghitung koefisien korelasi antara setiap pasangan variabel, yang kemudian divisualisasikan dalam bentuk matriks warna. Nilai koefisien berkisar antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan korelasi positif yang kuat, nilai mendekati -1 menunjukkan korelasi negatif yang kuat, dan nilai mendekati 0 menunjukkan tidak adanya korelasi (Leni, 2023).

2.7 Hyperparameter

Hyperparameter merupakan parameter yang nilainya ditetapkan sebelum proses pembelajaran dimulai, berbeda dengan parameter model yang dipelajari selama proses pelatihan. Pemilihan *hyperparameter* yang tepat penting dalam menentukan performa model jaringan saraf. Inisialisasi *hyperparameter* yang tepat dapat membantu model mencapai solusi optimal lokal (Uzun et al., 2024).

2.7.1 Hidden Layer dan Nodes

Hidden layer merupakan lapisan tersembunyi yang berada di antara *input layer* dan *output layer* pada arsitektur *neural network*. Menurut penelitian, jumlah *hidden layer* yang optimal bergantung pada kompleksitas permasalahan yang dihadapi. Semakin dalam arsitektur (lebih banyak *hidden layer*), semakin baik kemampuan model dalam mempelajari fitur-fitur kompleks, namun juga membutuhkan lebih banyak data training dan komputasi. Sementara itu jumlah *nodes* atau neuron pada setiap layer menentukan kapasitas model dalam mempelajari pola data. Pemilihan jumlah neuron yang terlalu sedikit dapat menyebabkan *underfitting*, sementara terlalu banyak neuron dapat mengakibatkan *overfitting* (Uzun et al., 2024). Dalam konteks LSTM, jumlah neuron ini juga menentukan dimensi *cell state* dan *hidden state*.

2.7.2 Epoch

Epoch merepresentasikan berapa kali keseluruhan dataset training diproses oleh model. Jumlah *epoch* optimal perlu mempertimbangkan *trade-off* antara waktu training dan konvergensi model. Terlalu sedikit *epoch* dapat menyebabkan model belum mencapai konvergensi, sementara terlalu banyak *epoch* dapat mengakibatkan *overfitting*.

2.7.3 Fungsi Optimasi

Optimizer menentukan bagaimana model diperbarui berdasarkan *loss function*. Fungsi optimasi memiliki peran penting dalam meningkatkan efisiensi pembelajaran dan akurasi model prediksi. Berbagai teknik optimasi, seperti Adam dan RMSProp dirancang untuk menyesuaikan bobot model secara adaptif demi meminimalkan kesalahan prediksi (Kristiyanti et al., 2024).

2.7.4 *Learning rate*

Learning rate mengontrol seberapa besar langkah *update* parameter dalam proses optimisasi. *Learning rate* yang terlalu besar dapat menyebabkan divergensi, sementara *learning rate* yang terlalu kecil membuat proses pembelajaran lambat (Huang et al., 2023). Pustaka *ReduceLROnPlateau* dapat digunakan untuk mengatasi masalah jumlah *learning rate* sehingga dapat mengoptimalkan proses pelatihan model (He et al., 2024).

2.7.5 *Timestep*

Dalam konteks LSTM, *timestep* menentukan panjang sekuens input yang diproses dalam satu waktu. Pemilihan *timestep* yang tepat sangat penting untuk menangkap dependensi temporal dalam data. *Timestep* yang terlalu pendek mungkin tidak dapat menangkap pola jangka panjang, sementara *timestep* yang terlalu panjang dapat meningkatkan kompleksitas komputasi (Furizal et al., 2024).

2.7.6 *Batch size*

Batch size menentukan jumlah sampel yang diproses dalam satu iterasi training. *Batch size* yang lebih besar dapat mempercepat *training* dan memberikan estimasi gradien yang lebih stabil, namun membutuhkan lebih banyak memori. *Batch size* yang umum digunakan berkisar antara 32 hingga 256 tergantung pada jumlah data yang digunakan.

2.7.7 *Fungsi Aktivasi*

Fungsi aktivasi memperkenalkan non-linearitas ke dalam model. Untuk LSTM, beberapa fungsi aktivasi yang umum digunakan adalah:

a. Sigmoid

Fungsi aktivasi *sigmoid* berfungsi untuk menghasilkan output antara 0 dan 1 sehingga digunakan pada *input* dan *output gate* dalam LSTM karena merepresentasikan probabilitas. *Sigmoid* efektif untuk kontrol informasi dalam *gates* LSTM karena menghasilkan nilai biner yang menentukan informasi diteruskan atau dilupakan.

b. Tanh

Fungsi aktivasi *tanh* menghasilkan output antara -1 dan 1 sehingga digunakan untuk *cell state update* dalam LSTM. *Tanh* membantu mengatasi masalah *vanishing gradient* lebih baik dibanding *sigmoid*.

c. ReLU

Fungsi aktivasi *Rectified Linear Unit (ReLU)* memiliki kemampuan untuk menangani input negatif lebih efektif (Dubey et al., 2022). ReLU dapat mempercepat konvergensi dan mengatasi masalah *vanishing gradient*.

2.7.8 Dropout

Dropout adalah teknik regularisasi yang sering digunakan dalam model deep learning untuk mencegah *overfitting*. Teknik ini bekerja dengan menonaktifkan neuron pada suatu layer secara acak dengan probabilitas tertentu selama proses pelatihan model, sehingga neuron tersebut tidak berkontribusi pada aktivasi layer berikutnya dan tidak mengalami pembaruan bobot selama *backpropagation* (Bérchez-Moreno et al., 2024). Nilai *dropout* yang umum berkisar antara 0.2 hingga 0.5.

2.8 Group K-Fold Cross Validation

Teknik validasi *k-fold cross validation* membagi dataset menjadi k bagian, di mana k-1 bagian digunakan untuk *training* dan 1 bagian untuk validasi, dengan proses diulang k kali. *K-fold cross validation* dapat digunakan untuk mendapatkan estimasi performa model yang lebih *robust*. Teknik ini sangat bermanfaat, terutama pada dataset dengan ukuran terbatas, karena memastikan setiap data digunakan untuk pelatihan dan pengujian. Salah satu pengembangan dari teknik validasi silang ini adalah *group k-fold cross validation*. Metode *group k-fold* melakukan validasi berdasarkan kategori yang tersedia pada data yang digunakan sehingga validasi akan dilakukan sesuai dengan jumlah kategori dalam data (Abhigyan, 2021).

2.9 Metrik Evaluasi

2.9.1 Mean Absolute Error (MAE)

MAE digunakan untuk menghitung rata-rata absolut dari selisih antara nilai prediksi dan nilai aktual tanpa memperhitungkan *outlier*. Semakin kecil hasil evaluasi MAE maka dapat dikatakan bahwa performa model prediksi semakin baik (Xiao et al., 2021).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{pred} - y_{act}| \dots \dots \dots (13)$$

Keterangan:

n : jumlah baris data

y_{pred} : nilai prediksi target

y_{act} : nilai aktual target

2.9.2 R-squared (R^2)

R-squared digunakan untuk menunjukkan seberapa baik model prediksi bekerja dengan data. Nilai R^2 berada dalam rentang antara nol dan satu dimana ketika nilai R^2 semakin mendekati satu menandakan model yang lebih baik (Plevris et al., 2022).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{act} - y_{pred})^2}{\sum_{i=1}^n (y_{act} - y_{avg})^2} \dots \dots \dots (14)$$

Keterangan:

n : jumlah baris data

y_{pred} : nilai target hasil prediksi

y_{act} : nilai aktual target

y_{avg} : rata-rata nilai aktual target

BAB 3

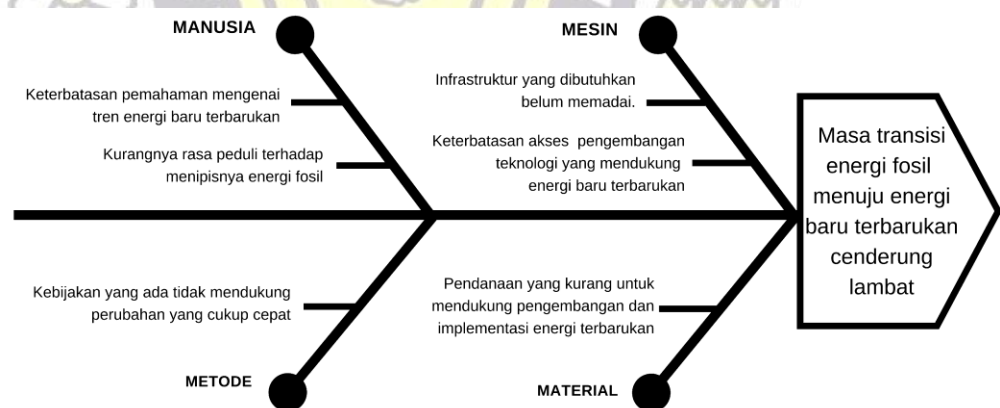
ANALISIS DAN PERANCANGAN SISTEM

3.1 Analisis Sistem

Tahap analisis sistem dalam sebuah penelitian perlu dilakukan dengan tujuan agar dapat mengetahui kebutuhan yang diperlukan saat memasuki proses perancangan sistem. Dengan analisis sistem yang menyeluruh, peneliti dapat mengenali aspek-aspek penting yang memengaruhi perancangan dan pengembangan sistem agar lebih tepat sasaran. Proses analisis sistem dalam penelitian ini terdiri dari analisis masalah dan analisis kebutuhan.

3.1.1 Analisis Masalah

Penelitian diawali dengan menemukan isu yang dapat diangkat melalui tahap analisis masalah. Analisis masalah dapat dilakukan untuk mengetahui sebab dan akibat dari suatu masalah. Teknik analisis masalah yang digunakan dalam penelitian ini adalah *fishbone diagram* yang mencakup empat kategori yaitu manusia, metode, material dan mesin atau sistem.



Gambar 5 *Fishbone* diagram

Fishbone diagram dengan empat faktor digunakan dengan tujuan agar penelitian berfokus pada masalah utama yang dibahas dalam penelitian. Penelitian ini memiliki fokus utama untuk menyelesaikan masalah yang tertera dalam *fishbone* diagram pada faktor manusia yaitu keterbatasan pemahaman tentang tren konsumsi energi baru terbarukan dan faktor mesin yaitu keterbatasan akses pengembangan teknologi.

3.1.2 Analisis Kebutuhan

Tahap analisis kebutuhan merupakan proses identifikasi untuk mengetahui hal-hal penting yang diperlukan dalam perancangan dan pengembangan sistem. Pada penelitian ini, analisis kebutuhan dibagi menjadi dua jenis, yaitu analisis kebutuhan fungsional dan non-fungsional.

1. Kebutuhan fungsional

Kebutuhan fungsional merupakan mencakup fitur-fitur yang harus dimiliki sistem agar dapat beroperasi sesuai tujuan. Adapun kebutuhan fungsional dalam penelitian ini adalah sebagai berikut:

- a. Sistem mampu membaca dan mengolah nilai-nilai yang diberikan oleh pengguna.
- b. Sistem mampu menentukan nilai prediksi konsumsi energi baru terbarukan berdasarkan nilai masukan yang diberi pengguna.
- c. Sistem mampu menampilkan grafik tren konsumsi energi baru terbarukan yang dapat dipahami oleh pengguna.

2. Kebutuhan non-fungsional

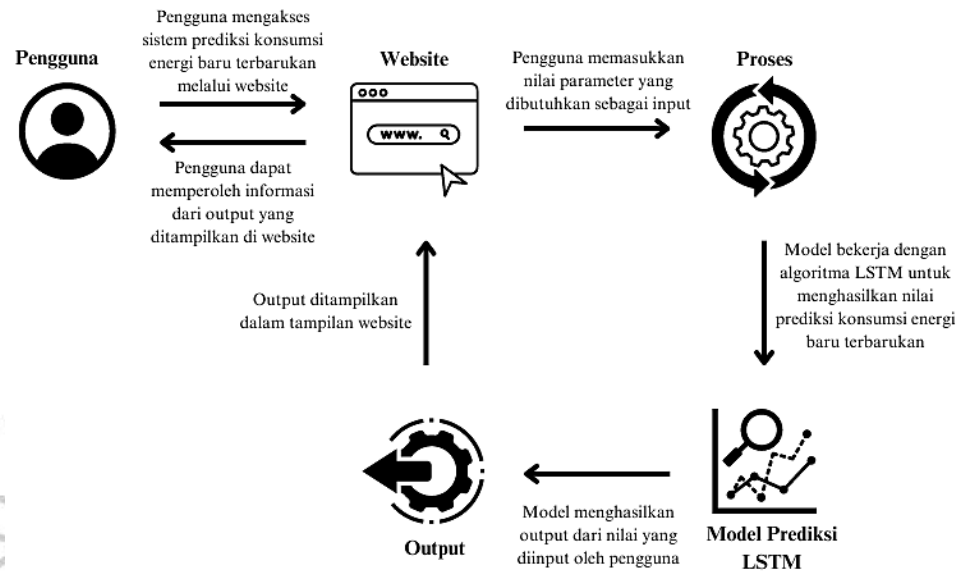
Kebutuhan non-fungsional mencakup aspek yang berkaitan dengan kinerja sistem. Adapun kebutuhan non-fungsional dalam penelitian ini adalah sebagai berikut:

- a. Sistem dibangun dengan desain antarmuka yang informatif dan interaktif.
- b. Aplikasi harus dapat diakses dari semua sistem operasi yang perangkatnya telah terhubung dengan jaringan internet.
- c. Sistem harus memiliki waktu pemrosesan yang efisien, sehingga prediksi dapat dilakukan dalam waktu yang wajar.

3.2 Pemodelan Sistem

Pemodelan sistem dilakukan untuk memberikan gambaran umum tentang sistem yang akan dibangun. Tahap ini menjadi panduan agar sistem yang dibangun memenuhi kebutuhan yang telah diidentifikasi pada proses sebelumnya.

3.2.1 Arsitektur Sistem



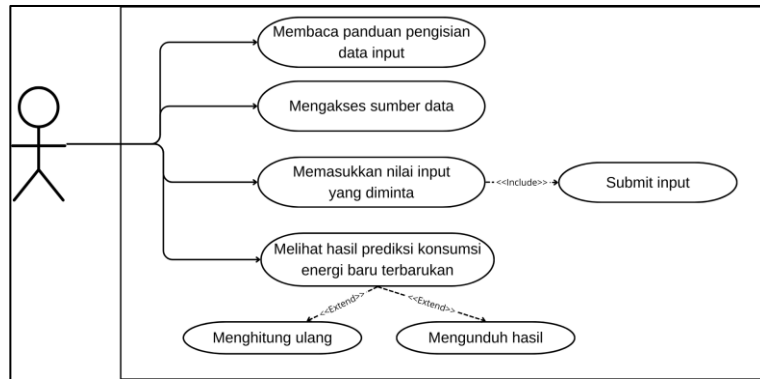
Gambar 6 Arsitektur sistem

Sistem prediksi konsumsi energi baru terbarukan berbasis algoritma *Long Short-Term Memory* (LSTM) dimulai ketika pengguna mengakses sistem melalui sebuah website. Di website tersebut, pengguna memasukkan nilai parameter yang dibutuhkan sebagai input untuk proses prediksi. Setelah data input diterima, sistem memulai proses pengolahan data menggunakan algoritma LSTM yang telah dilatih sebelumnya. Algoritma ini bekerja dengan menganalisis pola dari data yang dimasukkan oleh pengguna untuk menghasilkan prediksi konsumsi energi baru terbarukan.

Hasil dari pemrosesan data ini kemudian diolah oleh model prediksi LSTM dan menghasilkan output berupa nilai prediksi konsumsi energi. Output tersebut selanjutnya ditampilkan di tampilan website, sehingga pengguna dapat melihat hasil prediksi dengan jelas. Akhirnya, pengguna dapat menggunakan informasi yang disajikan pada website untuk analisis lebih lanjut atau untuk pengambilan keputusan terkait konsumsi energi baru terbarukan. Sistem ini memudahkan interaksi antara pengguna dan model prediksi melalui antarmuka web yang sederhana dan efektif.

3.2.2 Use Case Diagram

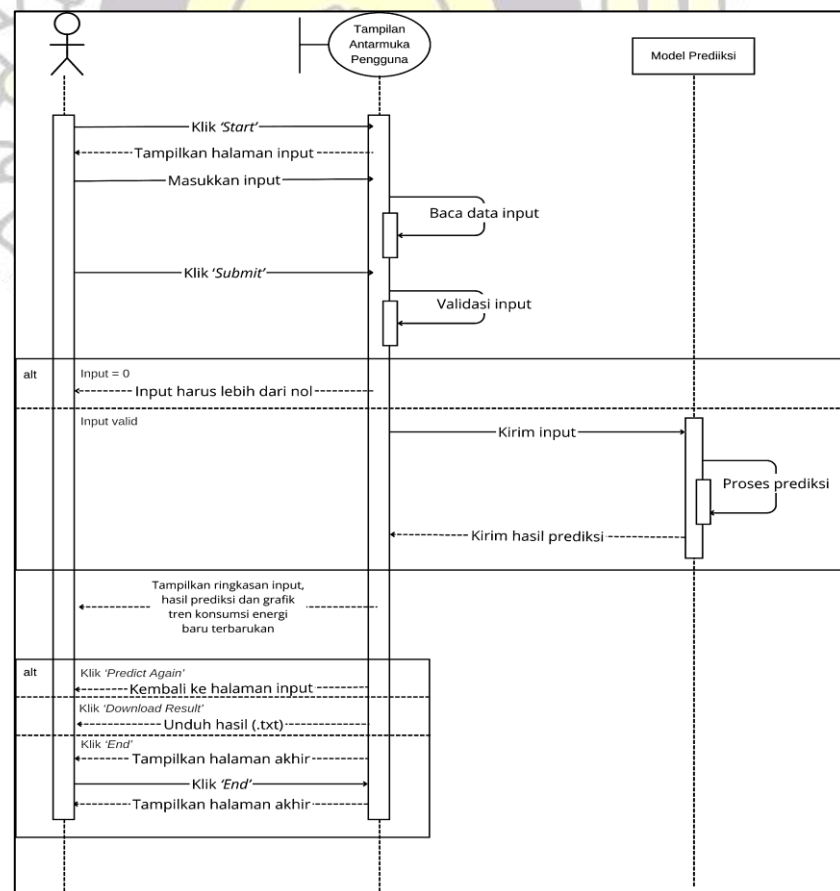
Use case diagram disusun untuk mengetahui apa saja interaksi yang dapat dilakukan pengguna kepada sistem. Diagram ini memiliki rincian tentang fungsionalitas sistem dan bagaimana sistem berfungsi menurut sudut pandang pengguna.



Gambar 7 Use case diagram

3.2.3 Sequence Diagram

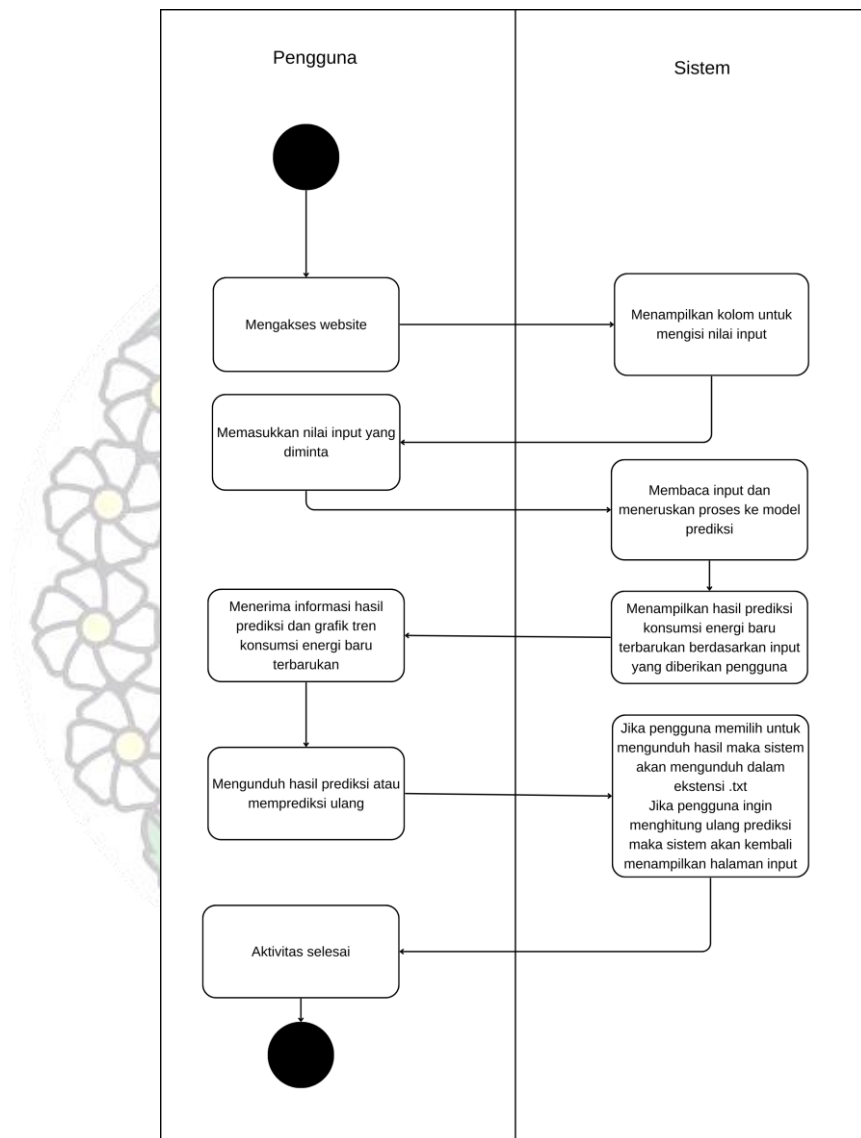
Sequence diagram dirangkai untuk memahami skenario interaksi antara pengguna, tampilan antarmuka dan model prediksi. Diagram ini akan membantu memastikan kelancaran alur sistem.



Gambar 8 Sequence diagram

3.2.4 Activity Diagram

Activity diagram dibuat agar setiap langkah dalam suatu alur kerja tergambar dengan jelas. Gambar berikut menjelaskan bagaimana pengguna dapat memperoleh informasi nilai prediksi konsumsi energi baru terbarukan.



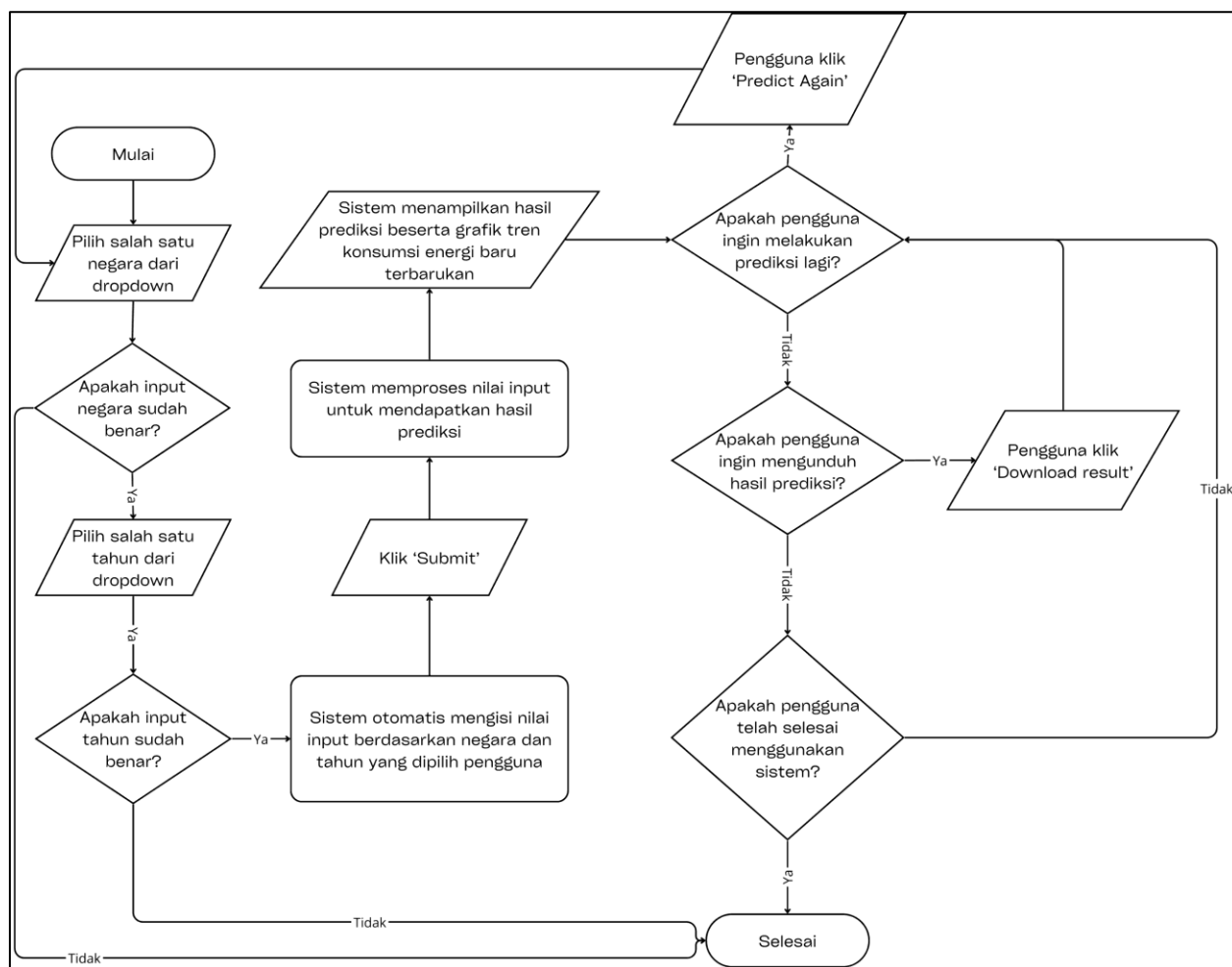
Gambar 9 Activity diagram

3.2.5 Flowchart

Flowchart (diagram alir) digunakan untuk memvisualisasikan alur proses dalam mencapai suatu tujuan. Pada penelitian ini, *flowchart* dibuat menggunakan simbol standar yang digunakan dalam pembuatan *flowchart*.

1. *Flowchart* sistem

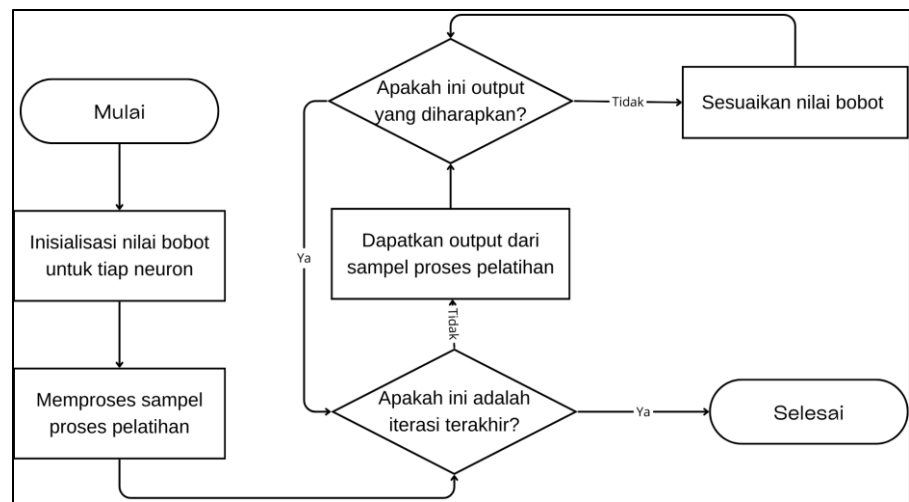
Diagram alur sistem yang dibangun adalah sebagai berikut:



Gambar 10 *Flowchart* sistem

2. Flowchart algoritma LSTM

Diagram alur algoritma yang digunakan adalah sebagai berikut:



Gambar 11 Flowchart algoritma LSTM

3.3 Analisis Proses

Analisis proses melibatkan identifikasi langkah-langkah yang ada dalam suatu alur kerja, serta pengukuran kinerja masing-masing bagian untuk menemukan atau area yang memerlukan perbaikan. Dalam tahap ini data dikumpulkan untuk mengidentifikasi variabel-variabel yang memengaruhi kinerja sistem.

3.3.1 Pengumpulan Data

Dalam penelitian ini data yang digunakan adalah data konsumsi energi tahunan pada negara berkembang di Asia. Data yang digunakan bersumber dari *World Development Indicator (WDI)* dan *Our World in Data (OWiD)*. Total data yang digunakan berjumlah 700 baris dan 8 kolom dengan rincian sebagai berikut:

Tabel 1 Rincian dataset

| No. | Kolom | Keterangan | Deskripsi | Tipe Data | Sumber Data |
|-----|--------------------------------------|--|---|---------------|-------------|
| 1. | <i>Country</i> | Negara | Negara berkembang di Asia termasuk Bangladesh, Brunei Darussalam, China, Filipina, India, Indonesia, Iran, Kamboja, Laos, Malaysia, Myanmar, Pakistan, Thailand dan Vietnam | <i>String</i> | WDI |
| 2. | <i>Year</i> | Tahun | 1974 - 2023 | Numerik | WDI |
| 3. | <i>GDP (\$)</i> | <i>Gross Domestic Product (GDP)</i> | Pendapatan bruto perkapita dalam USD | Numerik | WDI |
| 4. | <i>FDI_in (\$)</i> | <i>Foreign Direct Investment (FDI)</i> | Arus masuk investasi asing dalam USD | Numerik | WDI |
| 5. | <i>CO2_emission (ton per capita)</i> | <i>CO2 Emission</i> | Jumlah emisi karbon dalam ton perkapita | Numerik | OWiD |
| 6. | <i>FF_cons (kWh per person)</i> | <i>Fossil Fuel Consumption</i> | Jumlah konsumsi bahan bakar fosil dalam kilo-watt jam perkapita | Numerik | OWiD |
| 7. | <i>Energy use (kWh per person)</i> | <i>Energy Use</i> | Jumlah energi yang digunakan dalam kilo-watt jam perkapita | Numerik | OWiD |
| 8. | <i>RE_cons (kWh per person)</i> | <i>Renewable Energy Consumption</i> | Jumlah konsumsi energi baru terbarukan dalam kilo-watt jam perkapita | Numerik | OWiD |

3.3.2 Perhitungan Prediksi Konsumsi Energi Baru Terbarukan dengan Algoritma LSTM

1. Prapemrosesan Data

Tahap prapemrosesan data sangat mempengaruhi hasil akhir model.

Prapemrosesan data memastikan bahwa model mempelajari data dan menghasilkan model yang optimal.

a. Sampel data

Berdasarkan Tabel 1, berikut merupakan beberapa sampel data yang digunakan dalam penelitian ini:

Tabel 2 Sampel data

| <i>Country</i> | <i>Year</i> | <i>GDP</i> | <i>FDI_in</i> | <i>CO2_emission</i> | <i>FF_cons</i> | <i>Energy use</i> | <i>RE_cons</i> |
|----------------|-------------|------------|---------------|---------------------|----------------|-------------------|----------------|
| Indonesia | 1974 | 2016,68 | 182120,00 | 0,80 | 980,95 | 1020,78 | 39,83 |
| Indonesia | 1975 | 2321,71 | 129205,00 | 0,82 | 1178,08 | 1218,81 | 40,73 |
| Indonesia | 1976 | 2770,51 | 747590,00 | 0,92 | 1225,32 | 1248,63 | 23,32 |
| Indonesia | 1977 | 3322,82 | 235000,00 | 1,20 | 1573,11 | 1610,93 | 37,82 |
| Indonesia | 1978 | 3642,86 | 417709,00 | 1,33 | 1742,51 | 1768,87 | 26,37 |
| Indonesia | 1979 | 3552,36 | 226000,00 | 1,32 | 1937,35 | 1965,98 | 28,63 |
| Indonesia | 1980 | 4891,60 | 180000,00 | 1,28 | 2037,16 | 2062,47 | 25,31 |

b. Penanganan outliers

Outliers merupakan nilai yang jauh berbeda dari sebagian besar data yang disebabkan oleh kesalahan pengukuran atau kejadian yang sangat jarang. *Outliers* dapat memengaruhi hasil analisis dan model. Penanganan *outliers* dalam penelitian ini dilakukan dengan *Interquartile Range (IQR)*. *IQR* dihitung sebagai selisih antara kuartil ketiga (Q3) dan kuartil pertama (Q1). *Outlier* didefinisikan sebagai data yang berada diluar rentang.

- Hitung Q1 dan Q3

Merujuk pada Tabel 2, diketahui bahwa jumlah sampel data (n) = 7, oleh karena itu dapat dihitung:

$$Q1 = 0,25 \times (n + 1) = 0,25 \times (7 + 1) = 2$$

$$Q3 = 0,75 \times (n + 1) = 0,75 \times (7 + 1) = 6$$

Maka dapat diketahui bahwa dengan jumlah sampel data (n) = 7, Q1 berada pada data ke-2 dan Q3 berada pada data ke-6.

- Hitung transformasi $\log(x + 1)$

Tahap ini diperlukan untuk membantu mengurangi *skewness* (kemiringan) distribusi data, mengatasi *outlier* yang ekstrem, dan membuat perhitungan IQR lebih stabil dan akurat. Tabel berikut menunjukkan transformasi log yang dihitung berdasarkan sampel data pada Tabel 2.

Tabel 3 Transformasi log

| <i>GDP</i> | <i>FDI_in</i> | <i>CO2_emission</i> | <i>FF_cons</i> | <i>Energy use</i> | <i>RE_cons</i> |
|------------|---------------|---------------------|----------------|-------------------|----------------|
| 7,61 | 12,11 | 0,59 | 6,89 | 6,93 | 3,71 |
| 7,75 | 11,77 | 0,60 | 7,07 | 7,11 | 3,73 |
| 7,93 | 13,52 | 0,65 | 7,11 | 7,13 | 3,19 |
| 8,11 | 12,37 | 0,79 | 7,36 | 7,39 | 3,66 |
| 8,20 | 12,94 | 0,85 | 7,46 | 7,48 | 3,31 |
| 8,18 | 12,33 | 0,84 | 7,57 | 7,58 | 3,39 |
| 8,50 | 12,10 | 0,82 | 7,62 | 7,63 | 3,27 |

- Penanganan *outlier* untuk kolom *GDP* dengan teknik

$$GDP = [7.61, 7.75, 7.93, 8.11, 8.18, 8.20, 8.50]$$

$$Q1 = 7,75$$

$$Q3 = 8,20$$

$$IQR = 8,20 - 7,75 = 0,45$$

$$Lower\ bound = Q1 - 1,5 \times IQR = 7,75 - 1,5 \times 0,45 = 7,07$$

$$Upper\ bound = Q3 + 1,5 \times IQR = 8,20 + 1,5 \times 0,45 = 8,87$$

Berdasarkan *lower bound* dan *upper bound*, tidak ada *outlier* yang ditemukan pada kolom *GDP*. Oleh karena itu nilai dapat dikembalikan ke nilai semula dengan menggunakan rumus $\exp(x) - 1$.

- Penanganan *outlier* untuk kolom *FDI_in*

$$FDI_in = [11.77, 12.10, 12.11, 12.33, 12.37, 12.94, 13.52]$$

$$Q1 = 12,10$$

$$Q3 = 12,94$$

$$IQR = 12,94 - 12,10 = 0,84$$

$$Lower\ bound = Q1 - 1,5 \times IQR = 12,10 - 1,5 \times 0,84 = 10,84$$

$$Upper\ bound = Q3 + 1,5 \times IQR = 12,94 + 1,5 \times 0,84 = 14,20$$

Berdasarkan *lower bound* dan *upper bound*, tidak ada *outlier* yang ditemukan pada kolom *FDI_in*. Oleh karena itu nilai dapat dikembalikan ke nilai semula dengan menggunakan rumus $\exp(x) - 1$.

- Penanganan *outlier* untuk kolom *CO2_emission*

$$CO2_emission = [0.59, 0.60, 0.65, 0.79, 0.82, 0.84, 0.85]$$

$$Q1 = 0,60$$

$$Q3 = 0,84$$

$$IQR = 0,84 - 0,60 = 0,24$$

$$Lower\ bound = Q1 - 1,5 \times IQR = 0,60 - 1,5 \times 0,24 = 0,24$$

$$Upper\ bound = Q3 + 1,5 \times IQR = 0,84 + 1,5 \times 0,24 = 1,20$$

Berdasarkan *lower bound* dan *upper bound*, tidak ada *outlier* yang ditemukan pada kolom *CO2_emission*. Oleh karena itu nilai dapat dikembalikan ke nilai semula dengan menggunakan rumus $\exp(x) - 1$.

- Penanganan *outlier* untuk kolom *FF_cons*

$$FF_cons = [6.89, 7.07, 7.11, 7.36, 7.46, 7.57, 7.62]$$

$$Q1 = 7,07$$

$$Q3 = 7,57$$

$$IQR = 7,57 - 7,07 = 0,5$$

$$Lower\ bound = Q1 - 1,5 \times IQR = 7,07 - 1,5 \times 0,5 = 6,32$$

$$Upper\ bound = Q3 + 1,5 \times IQR = 7,57 + 1,5 \times 0,5 = 8,29$$

Berdasarkan *lower bound* dan *upper bound*, tidak ada *outlier* yang ditemukan pada kolom *FF_cons*. Oleh karena itu nilai dapat dikembalikan ke nilai semula dengan menggunakan rumus $\exp(x) - 1$.

- Penanganan *outlier* untuk kolom *Energy use*

$$\text{Energy use} = [6.93, 7.11, 7.13, 7.39, 7.48, 7.58, 7.63]$$

$$Q1 = 7,11$$

$$Q3 = 7,58$$

$$IQR = 7,58 - 7,11 = 0,47$$

$$\text{Lower bound} = Q1 - 1,5 \times IQR = 7,11 - 1,5 \times 0,47 = 6,41$$

$$\text{Upper bound} = Q3 + 1,5 \times IQR = 7,58 + 1,5 \times 0,47 = 8,28$$

Berdasarkan *lower bound* dan *upper bound*, tidak ada *outlier* yang ditemukan pada kolom *Energy use*. Oleh karena itu nilai dapat dikembalikan ke nilai semula dengan menggunakan rumus $\exp(x) - 1$.

- Penanganan *outlier* untuk kolom *RE_cons*

$$\text{RE_cons} = [3.19, 3.27, 3.31, 3.39, 3.66, 3.71, 3.73]$$

$$Q1 = 3,27$$

$$Q3 = 3,71$$

$$IQR = 3,71 - 3,27 = 0,44$$

$$\text{Lower bound} = Q1 - 1,5 \times IQR = 3,27 - 1,5 \times 0,44 = 2,61$$

$$\text{Upper bound} = Q3 + 1,5 \times IQR = 3,71 + 1,5 \times 0,44 = 4,37$$

Berdasarkan *lower bound* dan *upper bound*, tidak ada *outlier* yang ditemukan pada kolom *RE_cons*. Oleh karena itu nilai dapat dikembalikan ke nilai semula dengan menggunakan rumus $\exp(x) - 1$.

c. Penanganan *missing values*

Data yang hilang (*missing values*) dalam penelitian ini terjadi karena tidak tersedianya data khususnya pada tahun lampau. Brunei Darussalam, Kamboja dan Myanmar menjadi tiga negara yang memiliki *missing values* pada kolom *RE_cons* atau dalam hal ini adalah kolom yang menjadi target prediksi. Tantangan lainnya yaitu terkait *missing value* yang terletak pada baris pertama. Hal ini menjadikan metode interpolasi linier dan *forward fill* tidak bisa digunakan. Oleh karena itu diperlukan metode yang sesuai untuk mengisi nilai yang hilang tersebut sebelum dilakukan prediksi konsumsi energi baru terbarukan dengan algoritma LSTM.

Pada penelitian ini, metode *backward fill* diterapkan untuk mengatasi *missing value*. Berikut ini adalah perbandingan antara data yang masih mengandung *missing value* dan data yang telah diisi menggunakan *backward fill*:

Tabel 4 Penanganan *missing value* dengan *backward fill*

| <i>Country</i> | <i>Year</i> | RE_cons | |
|----------------|-------------|---|---|
| | | Sebelum <i>backward fill</i> | Sesudah <i>backward fill</i> |
| Myanmar | 1974 | - | 18,80 |
| Myanmar | 1975 | - | 18,80 |
| Myanmar | 1976 | - | 18,80 |
| Myanmar | 1977 | - | 18,80 |
| Myanmar | 1978 | - | 18,80 |
| Myanmar | 1979 | - | 18,80 |
| Myanmar | 1980 | - | 18,80 |
| Myanmar | 1981 | - | 18,80 |
| Myanmar | 1982 | - | 18,80 |
| Myanmar | 1983 | - | 18,80 |
| Myanmar | 1984 | - | 18,80 |
| Myanmar | 1985 | - | 18,80 |
| Myanmar | 1986 | - | 18,80 |
| Myanmar | 1987 | - | 18,80 |
| Myanmar | 1988 | - | 18,80 |
| Myanmar | 1989 | - | 18,80 |
| Myanmar | 1990 | 18,80 | 18,80 |

d. Normalisasi

Normalisasi data merupakan proses yang bertujuan untuk menyamakan skala data. Penelitian ini menggunakan *MinMax scaling* yang sesuai untuk data deret waktu. Berikut adalah sampel data yang telah dinormalisasi menggunakan *MinMax scaling*:

Tabel 5 Normalisasi data

| <i>GDP</i> | <i>FDI_in</i> | <i>CO2_emission</i> | <i>FF_cons</i> | <i>Energy use</i> | <i>RE_cons</i> |
|-------------------|----------------------|----------------------------|-----------------------|--------------------------|-----------------------|
| 0,00 | 0,09 | 0,00 | 0,00 | 0,00 | 0,95 |
| 0,11 | 0,00 | 0,04 | 0,19 | 0,19 | 1,00 |
| 0,26 | 1,00 | 0,22 | 0,23 | 0,22 | 0,00 |
| 0,45 | 0,17 | 0,75 | 0,56 | 0,57 | 0,83 |
| 0,57 | 0,47 | 1,00 | 0,72 | 0,72 | 0,18 |
| 0,53 | 0,16 | 0,97 | 0,91 | 0,91 | 0,31 |
| 1,00 | 0,08 | 0,91 | 1,00 | 1,00 | 0,11 |

2. Perhitungan Prediksi Konsumsi Energi Baru Terbarukan

Tabel 6 Nilai input, bobot dan bias

| Variabel | Timestep (t) = 0 | Timestep (t) = 1 |
|-----------------|-------------------------|-------------------------|
| X_t | 0,00000 | 0,10610 |
| W_f | -8,55165 | -1,25693 |
| W_i | -1,49224 | 2,84737 |
| W_c | 7,43861 | -1,03959 |
| W_o | 2,56026 | -1,13412 |
| b_f | -2,57430 | 8,80652 |
| b_i | -1,57454 | -4,99221 |
| b_c | -3,41876 | -2,32592 |
| b_o | -4,53060 | 2,80204 |

Pada Tabel 6 indeks *timestep* 0 dan 1 yang digunakan pada perhitungan menggambarkan perhitungan yang terjadi untuk seluruh *timestep* yang digunakan dalam model prediksi. Selain itu, berikut juga diuraikan data yang perlu diketahui untuk menghitung

prediksi konsumsi energi baru terbarukan dengan algoritma LSTM. Berikut adalah keterangan variabel yang digunakan:

X_t : Input

W_f : Bobot yang digunakan saat *forget gate*

W_i : Bobot yang digunakan saat *input gate*

W_c : Bobot yang digunakan saat *cell state*

W_o : Bobot yang digunakan saat *output gate*

b_f : Bias yang digunakan saat *forget gate*

b_i : Bias yang digunakan saat *input gate*

b_c : Bias yang digunakan saat *cell state*

b_o : Bias yang digunakan saat *output gate*

t : *Time step*

σ : Fungsi aktivasi *sigmoid*

\tanh : Fungsi aktivasi *hyperbolic tangent*

Selanjutnya dapat dilakukan perhitungan prediksi konsumsi energi baru terbarukan dengan algoritma LSTM. Apabila ***time step* (t) = 0**, **$h_{t-1} = 0$** dan **$c_{t-1} = 0$** , maka dapat dilakukan perhitungan prediksi konsumsi energi baru terbarukan dengan algoritma LSTM yang dirincikan sebagai berikut:

- *Forget gate* (f_t)

$$\begin{aligned} f_t &= \sigma(W_f \cdot X_t + W_f \cdot h_{t-1} + b_f) \\ &= \sigma((-8,55165 \cdot 0) + (-8,55165 \cdot 0) + (-2,57430)) \\ &= \sigma(-2,57430) \\ &= 0,07081 \end{aligned}$$

- *Input gate* (i_t)

$$\begin{aligned} i_t &= \sigma(W_i \cdot X_t + W_i \cdot h_{t-1} + b_i) \\ &= \sigma((-1,49224 \cdot 0) + (-1,49224 \cdot 0) + (-1,57454)) \\ &= \sigma(-1,57454) \\ &= 0,17157 \end{aligned}$$

- *Candidate cell state* (\hat{C}_t)

$$\begin{aligned} \hat{C}_t &= \tanh(W_c \cdot X_t + W_c \cdot h_{t-1} + b_c) \\ &= \tanh((7,43861 \cdot 0) + (7,43861 \cdot 0) + (-3,41876)) \\ &= \tanh(-3,41876) \\ &= -0,99786 \end{aligned}$$

- *Cell state (c_t)*

$$\begin{aligned} c_t &= f_t \cdot c_{t-1} + i_t \dot{C}_t \\ &= (0,07081 \cdot 0) + (0,17157 \cdot -0,99786) \\ &= -0,17120 \end{aligned}$$

- *Output gate (O_t)*

$$\begin{aligned} O_t &= \sigma(W_o \cdot X_t + W_o \cdot h_{t-1} + b_o) \\ &= \sigma((2,56026 \cdot 0) + (2,56026 \cdot 0) + (-4,53060)) \\ &= \sigma(-4,53060) \\ &= 0,01066 \end{aligned}$$

- *Hidden state (H_t)*

$$\begin{aligned} H_t &= O_t \cdot \tanh(c_t) \\ &= 0,01066 \cdot \tanh(-0,17120) \\ &= -0,00181 \end{aligned}$$

Pada perhitungan *timestep* (t) = 0 didapatkan hasil $c_t = -0,17120$ dan $h_t = -0,00181$. Selanjutnya, kedua nilai ini akan menjadi pengaruh dalam perhitungan *timestep* (t) = 1.

Jika *timestep* (t) = 1, $h_{t-1} = -0,00181$ dan $c_{t-1} = -0,17120$, maka:

- *Forget gate (f_t)*

$$\begin{aligned} f_t &= \sigma(W_f \cdot X_t + W_f \cdot h_{t-1} + b_f) \\ &= \sigma((-1,25693 \cdot 0,10610) + (-1,25693 \cdot -0,00181) + 8,80652) \\ &= \sigma(8,67543) \\ &= 0,99983 \end{aligned}$$

- *Input gate (i_t)*

$$\begin{aligned} i_t &= \sigma(W_i \cdot X_t + W_i \cdot h_{t-1} + b_i) \\ &= \sigma((2,84737 \cdot 0,10610) + (2,84737 \cdot -0,00181) + (-4,99221)) \\ &= \sigma(-4,69525) \\ &= 0,00906 \end{aligned}$$

- *Candidate cell state (\dot{C}_t)*

$$\begin{aligned} \dot{C}_t &= \tanh(W_{\dot{C}} \cdot X_t + W_{\dot{C}} \cdot h_{t-1} + b_{\dot{C}}) \\ &= \tanh((-1,03959 \cdot 0,10610) + (-1,03959 \cdot -0,00181) + (-2,32592)) \\ &= \tanh(-2,43434) \\ &= -0,98475 \end{aligned}$$

- *Cell state (c_t)*

$$\begin{aligned} c_t &= f_t \cdot c_{t-1} + i_t \hat{c}_t \\ &= (0,99983 \cdot -0,17120) + (0,00906 \cdot -0,98475) \\ &= -0,18009 \end{aligned}$$

- *Output gate (O_t)*

$$\begin{aligned} O_t &= \sigma(W_o \cdot X_t + W_o \cdot h_{t-1} + b_o) \\ &= \sigma((-1,13412 \cdot 0,10610) + (-1,13412 \cdot -0,00181) + (2,80204)) \\ &= \sigma(2,68376) \\ &= 0,93606 \end{aligned}$$

- *Hidden state (H_t)*

$$\begin{aligned} H_t &= O_t \cdot \tanh(c_t) \\ &= 0,93606 \cdot \tanh(-1,8009) \\ &= -0,16678 \end{aligned}$$

Perhitungan ini terus dilakukan untuk setiap *timestep* dengan nilai c_t dan h_t yang terus diperbarui. Nilai c_t dan h_t yang dihasilkan pada langkah ini akan diteruskan pada langkah berikutnya sehingga model terus belajar untuk menghasilkan prediksi terbaik.

3. Denormalisasi

Denormalisasi data penting untuk dilakukan ketika data telah dinormalisasi sebelum pelatihan. Tahap denormalisasi bertujuan agar metrik evaluasi yang dihitung berada dalam skala sebenarnya.

$$X = X_{scaled} \times (X_{max} - X_{min}) + X_{min} \dots \dots \dots (15)$$

Keterangan:

X : nilai dalam skala sebenarnya

X_{scaled} : nilai setelah normalisasi

X_{max} : nilai maksimum dalam kolom

X_{min} : nilai minimum dalam kolom

3.4 Pseudocode

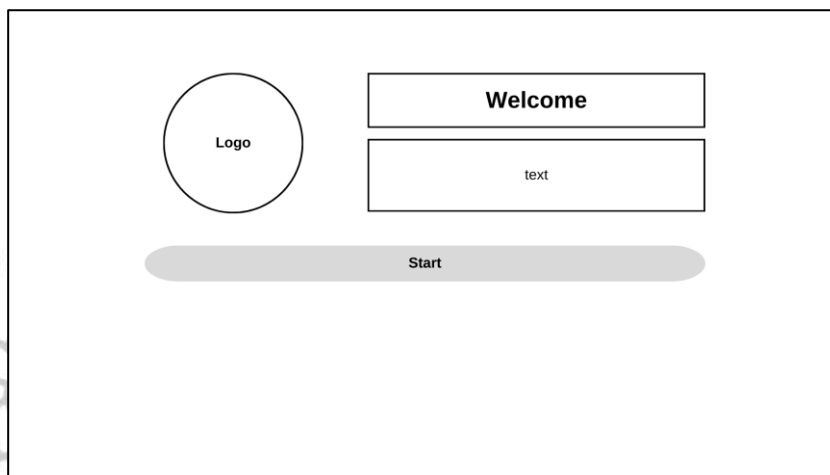
1. Load dataset
 - a. Load the dataset from 'dataset_14c_new.csv'
 - b. Display dataset info
2. Define feature columns and target column
 - a. `feature_columns = ['GDP ($)', 'FDI_in ($)', 'CO2_emission_percapita (ton per person)', 'Energyuse_percapita (kWh per person)', 'FF_cons (kWh per person)']`
 - b. `target_column = 'RE_cons (kWh per person)'`
3. Handle missing values using BFFill
 - a. Apply backward fill (BFFill) to handle missing values
4. Handle outliers using IQR and Winsorizing
 - a. For each column in `feature_columns` and `target_column`:
 - i. Detect outliers with IQR
 - ii. Apply Winsorizing to limit the extreme values
5. Normalize data using MinMaxScaler
 - a. Apply MinMaxScaler to scale the dataset between 0 and 1
6. Save the processed dataset
 - a. Save the normalized dataset as 'fulldata_14c.csv'
7. Display variables correlation
 - a. Show feature importance
 - b. Show a heatmap of feature correlations
8. Group K-Fold Cross Validation (14 folds)
 - a. Define 14-fold Group K-Fold cross-validation
 - b. For each fold:
 - i. Split dataset into training and testing sets based on category
 - ii. Build and train an LSTM model on the training set
 - iii. Evaluate the model on the test set
 - iv. Track performance metrics and save evaluation results
 - v. Select the best model based on performance
9. Testing and final evaluation with the best model
 - a. Use the best model to predict on the test set
 - b. Denormalize the predictions
 - c. Calculate performance metrics (MAE, R-squared) for the final model
10. Print evaluation results
 - a. Print the average MAE, R-squared, and the best model's MAE and R-squared
11. Visualize results
 - a. Display graphs for loss/validation loss and predicted vs actual values

3.5 Perancangan Antarmuka Pengguna

Perancangan antarmuka bertujuan untuk memberikan gambaran terkait tata letak sistem yang nantinya akan berinteraksi dengan pengguna. Rancangan antarmuka sistem pada penelitian ini adalah sebagai berikut:

3.5.1 Halaman awal

Halaman ini merupakan tampilan yang pertama kali dilihat oleh pengguna. Halaman awal ini berisi pengenalan tentang sistem dan juga tombol mulai untuk memasuki halaman input.



Gambar 12 Halaman awal

3.5.2 Halaman input

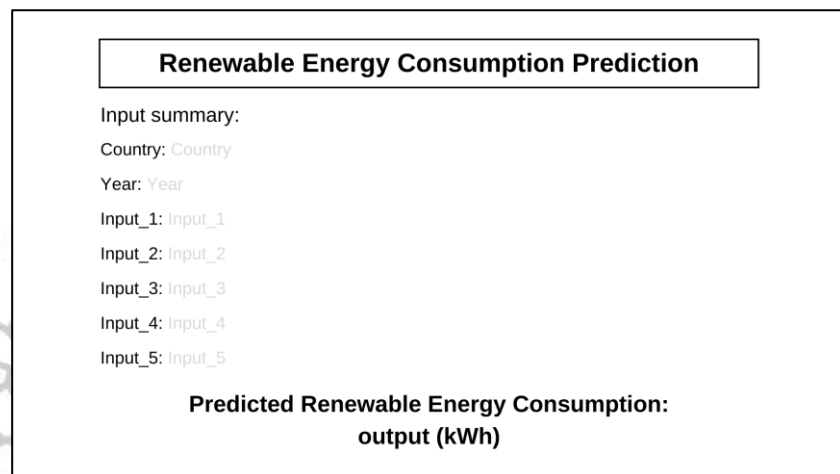
Halaman input merupakan halaman berikutnya yang akan tampil setelah pengguna memilih tombol mulai pada halaman awal. Halaman input meminta pengguna untuk memasukkan empat nilai numerik yaitu *Gross Domestic Products (GDP)*, *Foreign Direct Investment (FDI)*, *CO2 Emissions*, *Fossil Fuel Consumption* dan *Energy Use*.

 The image shows a wireframe of the input page. At the top, there is a header box labeled 'Input data'. Below it is a box for 'Data filling guidance'. Then, there are two drop-down menu boxes labeled 'Drop-down pilihan negara' and 'Drop-down pilihan tahun'. Below these are five input fields arranged in two columns. The left column has three fields labeled 'Input_1', 'Input_2', and 'Input_3'. The right column has two fields labeled 'Input_4' and 'Input_5'. Each input field has a red asterisk next to its label. At the bottom center, there is a wide, rounded rectangular button labeled 'Predict'.

Gambar 13 Halaman input

3.5.3 Halaman output

Halaman output menampilkan input yang sebelumnya telah diisi oleh pengguna dan juga hasil prediksi berupa nilai perkiraan konsumsi energi baru terbarukan dalam kilowatt-jam per orang. Halaman ini juga menyediakan tampilan grafik untuk menampilkan konsumsi energi baru terbarukan setiap tahunnya. Lalu pengguna bisa memilih tombol untuk menghitung prediksi lagi, mengunduh data ataupun menyelesaikan proses.



Renewable Energy Consumption Prediction

Input summary:

Country: Country

Year: Year

Input_1: Input_1

Input_2: Input_2

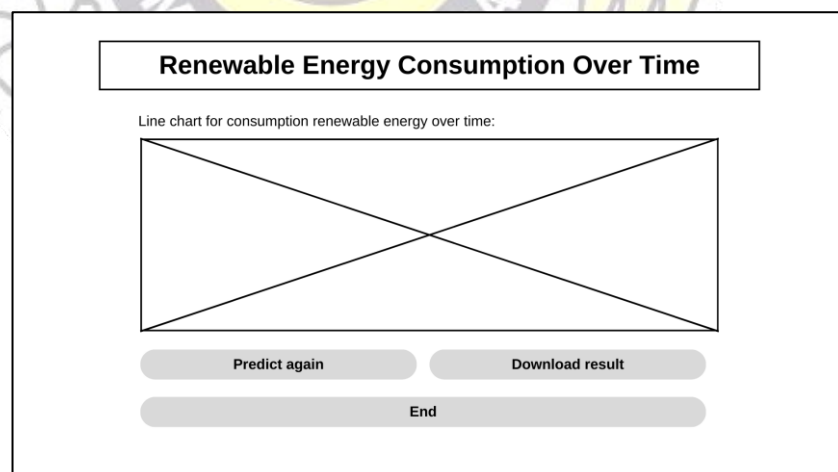
Input_3: Input_3

Input_4: Input_4

Input_5: Input_5

Predicted Renewable Energy Consumption:
output (kWh)

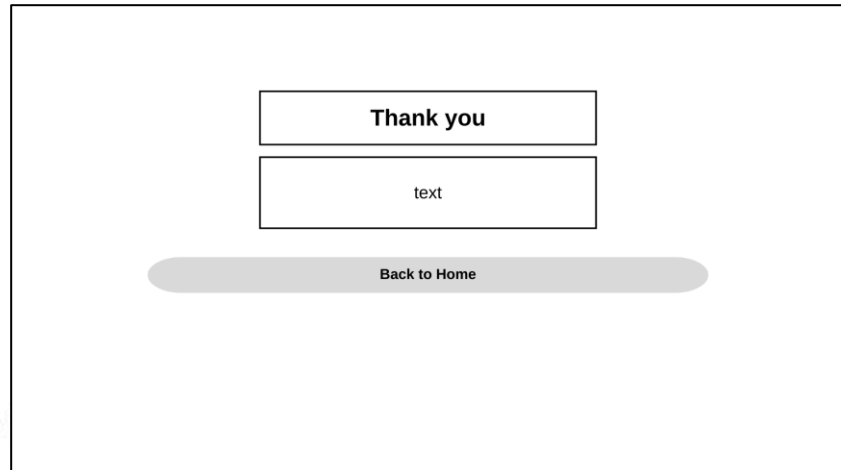
Gambar 14 Halaman output



Gambar 15 Halaman output grafik

3.5.4 Halaman akhir

Halaman terakhir merupakan halaman yang menampilkan teks ucapan terima kasih. Pada halaman ini juga terdapat tombol untuk memulai ulang web sehingga ketika pengguna klik tombol “*Back to Home*” maka ini akan mengarahkan pengguna ke halaman awal lagi.



Gambar 16 Halaman akhir



BAB 4

HASIL DAN PEMBAHASAN

4.1 Spesifikasi Perangkat Keras

Program dijalankan menggunakan server *Jupyter Notebook* melalui *Visual Studio Code* dengan spesifikasi perangkat sebagai berikut:

Tabel 7 Spesifikasi perangkat keras

| | |
|-----------|--|
| Merek | Acer Aspire A514-54 |
| Processor | 11 th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz (8 CPUs) ~2.4GHz |
| GPU | NVIDIA® GeForce® MX250 2GB |
| RAM | 16 GB |
| Disk SSD | 512 GB |

4.2 Hasil Prapemrosesan Data

Prapemrosesan data adalah langkah awal yang sangat penting untuk memastikan data dalam kondisi terbaik sebelum digunakan dalam proses pembelajaran model. Tahap ini bertujuan untuk membersihkan, menormalkan, dan mempersiapkan data agar model dapat mempelajarinya secara efektif. Kualitas data yang baik tidak hanya berkontribusi pada kinerja model yang optimal tetapi juga berperan penting dalam meningkatkan akurasi hasil prediksi.

Namun, dalam penelitian ini, keterbatasan ketersediaan data menjadi salah satu tantangan utama. Jumlah data yang terbatas dapat berdampak pada performa model, sehingga diperlukan pendekatan yang hati-hati dalam menangani data. Metode penanganan seperti penghapusan baris data kurang ideal untuk digunakan karena dapat semakin mengurangi jumlah data yang tersedia.

4.2.1 Hasil Penanganan *Outliers*

Outliers yang telah dideteksi oleh IQR kemudian akan ditangani dengan teknik *winsorizing* seperti pada pembahasan perhitungan sebelumnya. Teknik *winsorizing* ini akan mengubah nilai diluar

ambang batas tanpa mengubah rentang data secara keseluruhan. Hasil penanganan *outlier* dapat dilihat pada tabel berikut:

Tabel 8 Jumlah *outliers* yang ditangani

| Kolom | Jumlah <i>outlier</i> yang ditemukan | Jumlah <i>outlier</i> setelah <i>winsorizing</i> |
|---------------------|--------------------------------------|--|
| <i>GDP</i> | 5 | 0 |
| <i>FDI_in</i> | 18 | 0 |
| <i>CO2_emission</i> | 15 | 0 |
| <i>FF_cons</i> | 9 | 0 |
| <i>Energy use</i> | 8 | 0 |
| <i>RE_cons</i> | 47 | 0 |

Teknik *winsorizing* memastikan tidak ada nilai yang berada diluar rentang dengan cara ‘memotong’ *outliers*. Nilai yang lebih kecil dari ambang batas bawah akan diganti dengan ambang batas bawah itu sendiri. Begitu juga pada nilai yang lebih besar dari ambang batas atas akan diganti dengan ambang batas atas. Cara ini menjamin tidak ada nilai yang dihapus atau dibuang sehingga tetap mempertahankan jumlah data yang ada.

4.2.2 Hasil Penanganan *Missing Value*

Dalam penelitian ini, metode *backward fill* digunakan untuk mengisi *missing value* pada kolom target demi menghindari adanya penghapusan data. Adapun hasil penanganan *missing value* dalam penelitian ini dapat dijelaskan lebih lanjut dalam tabel berikut:

Tabel 9 Jumlah *missing value* yang ditangani

| Country | Year | Jumlah <i>missing value</i> pada kolom <i>RE_cons</i> | |
|-------------------|-------------|---|------------------------------|
| | | Sebelum <i>backward fill</i> | Sesudah <i>backward fill</i> |
| Brunei Darussalam | 1974 - 2002 | 29 | 0 |
| Cambodia | 1974 - 1989 | 16 | 0 |
| Myanmar | 1974 - 1989 | 16 | 0 |

4.2.3 Hasil Normalisasi

Setelah melalui tahap penanganan *outlier* dan *missing value*, data akan dinormalisasi untuk memastikan semua data berada dalam standar yang sama. *MinMax scaling* digunakan untuk menyamakan skala data sehingga berada diantara 0 dan 1. Adapun seluruh hasil dari tahap prapemrosesan dapat dilihat pada Tabel 10 berikut ini:

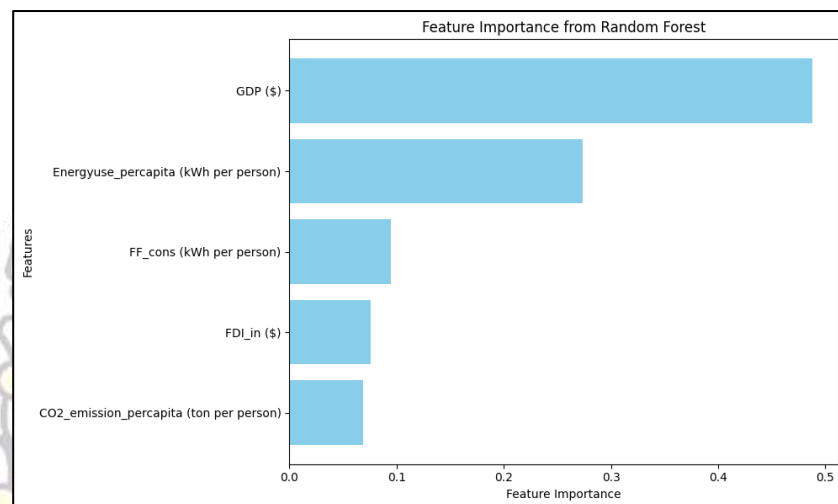
Tabel 10 Hasil prapemrosesan data

| <i>Country</i> | <i>GDP</i> | <i>FDI_in</i> | <i>CO2_emission</i> | <i>Energy use</i> | <i>FF_cons</i> | <i>RE_cons</i> |
|----------------|------------|---------------|---------------------|-------------------|----------------|----------------|
| Indonesia | 0,00000 | 0,00723 | 0,00000 | 0,00000 | 0,00000 | 0,01598 |
| Indonesia | 0,00644 | 0,05141 | 0,00476 | 0,02173 | 0,02419 | 0,01685 |
| Indonesia | 0,01591 | 0,02974 | 0,02554 | 0,02501 | 0,02999 | 0,00000 |
| Indonesia | 0,02756 | 0,00933 | 0,08571 | 0,06477 | 0,07267 | 0,01404 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| Iran | 0,23129 | 0,26738 | 0,93648 | 0,97451 | 0,97314 | 0,23129 |
| Iran | 0,41550 | 0,28392 | 0,96650 | 0,94114 | 0,95131 | 0,41550 |
| Iran | 0,49595 | 0,29886 | 0,98689 | 0,98550 | 0,99352 | 0,49595 |
| Iran | 0,47310 | 0,28339 | 1,00000 | 1,00000 | 1,00000 | 0,47310 |

Data dinormalisasi secara spesifik untuk setiap negara karena setiap negara memiliki pola distribusi data yang unik. Perbedaan ini dapat mempengaruhi akurasi dan konsistensi model dalam melakukan prediksi. Oleh karena itu, untuk memastikan bahwa model tetap dapat menghasilkan prediksi yang relevan dan akurat di setiap negara, normalisasi data dilakukan secara terpisah untuk masing-masing negara. Dengan cara ini, perbedaan dalam distribusi data antar negara dapat diminimalisir sehingga memungkinkan model untuk beradaptasi dengan karakteristik data masing-masing negara secara lebih efektif.

4.2.4 Pengecekan Korelasi Variabel

Pada penelitian ini divisualisasikan *feature importance ranking* menggunakan *Random Forest*. Pendekatan ini bertujuan untuk mengukur kontribusi setiap fitur terhadap target (Heidari et al., 2023). Visualisasi ini menghasilkan informasi tingkat pengaruh masing-masing fitur terhadap target prediksinya. *Feature importance* pada data yang digunakan dalam penelitian ini dapat divisualisasikan sebagai berikut:



Gambar 17 Tingkat pengaruh fitur terhadap target

Output:

Feature Importance:

GDP (\$): 0.4878

FDI_in (\$): 0.0755

CO2_emission_percapita (ton per person): 0.0685

Energyuse_percapita (kWh per person): 0.2736

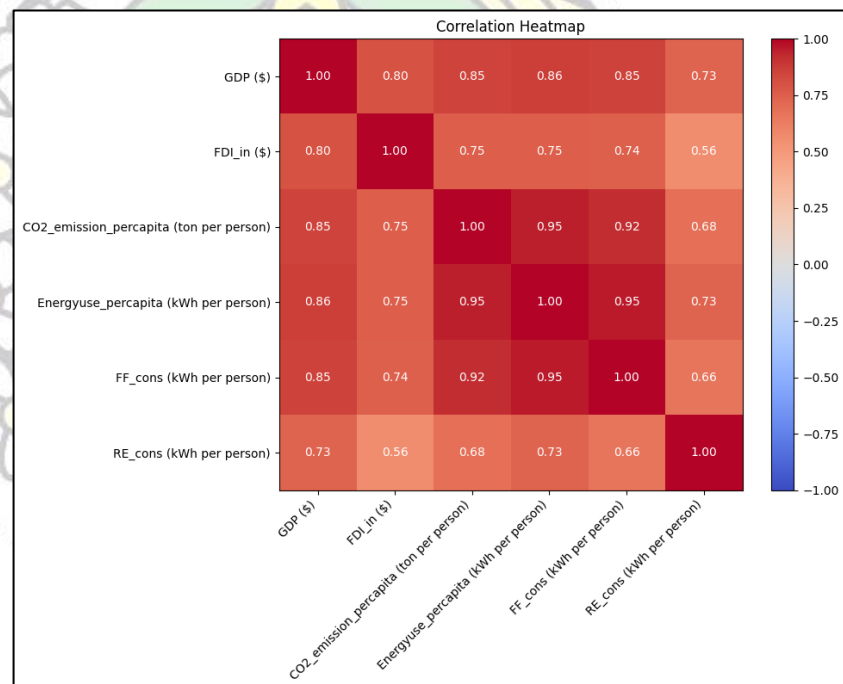
FF_cons (kWh per person): 0.0945

Total Feature Importance: 1.0

Hasil *feature importance* menunjukkan total nilai *importance* untuk seluruh fitur akan selalu mendekati satu karena telah dinormalisasi. Dari *output* di atas, dapat disimpulkan bahwa dari lima fitur yang dianalisis, *GDP (\$)* merupakan fitur yang paling mempengaruhi hasil prediksi konsumsi energi baru terbarukan. Hal ini menunjukkan bahwa variabel GDP memiliki kontribusi terbesar dalam mengurangi *impurity* (ketidakmurnian) di pohon-pohon pada model *Random*

Forest. Sebaliknya, fitur *CO2_emission_percapita (ton per person)* memberikan kontribusi paling sedikit dalam mempengaruhi hasil prediksi. Rendahnya *importance* dari *CO2_emission* menunjukkan bahwa fitur ini memiliki peran yang kecil dalam mengurangi *impurity* dibandingkan fitur lainnya. Interpretasi ini membantu dalam memahami keterkaitan dan hubungan setiap fitur terhadap target pada data yang digunakan dalam penelitian ini.

Selain itu, penelitian ini juga menggunakan *Heatmap Correlation* untuk mengukur hubungan antara setiap variabel dengan variabel lainnya. Hasil pengecekan korelasi variabel dalam penelitian ini dapat divisualisasikan pada gambar berikut:



Gambar 18 *Heatmap correlation*

Warna yang lebih hangat dalam korelasi dengan *heatmap* memiliki makna korelasi positif yang lebih kuat, sementara warna yang lebih dingin mendekati biru mengindikasikan korelasi negatif atau lemah. Berdasarkan Gambar 18, diketahui korelasi sebagai berikut:

- Pendapatan bruto perkapita (*GDP (\$)*) berkorelasi linier terhadap konsumsi energi baru terbarukan (*RE_cons (kWh per person)*) sebesar 0,73.

- Arus masuk investasi asing (FDI_{in} (\$)) berkorelasi linier terhadap konsumsi energi baru terbarukan (RE_{cons} (kWh per person)) sebesar 0,56.
- Emisi karbon ($CO2_{emissin}$ (ton per person)) berkorelasi linier terhadap konsumsi energi baru terbarukan (RE_{cons} (kWh per person)) sebesar 0,68.
- Jumlah energi yang digunakan ($Energy\ use$ (kWh per person)) berkorelasi linier terhadap konsumsi energi baru terbarukan (RE_{cons} (kWh per person)) sebesar 0,73.
- Jumlah bahan bakar fosil (FF_{cons} (kWh per person)) berkorelasi linier terhadap konsumsi energi baru terbarukan (RE_{cons} (kWh per person)) sebesar 0,66.

Berdasarkan analisis korelasi linier, terdapat hubungan positif antara konsumsi energi baru terbarukan (RE_{cons} (kWh per person)) dengan kelima variabel yang dianalisis. Secara keseluruhan, analisis ini mengindikasikan bahwa faktor ekonomi dan pola konsumsi energi memengaruhi tingkat penggunaan energi terbarukan, meskipun kekuatan hubungan-hubungan tersebut bervariasi.

4.3 Analisis Arsitektur Model

4.3.1 Eksplorasi *hyperparameter*

1. Jumlah *hidden layer* dan *nodes*

Uzun et al., dalam penelitiannya menyimpulkan bahwa jumlah *hidden layer* dan neuron yang optimal bergantung pada kompleksitas masalah yang dihadapi. Lebih banyak *hidden layer* meningkatkan kemampuan model dalam mempelajari data, namun membutuhkan data dan waktu komputasi lebih banyak. Selain itu, neuron yang terlalu sedikit dapat menghambat model dalam mencapai konvergensi sedangkan terlalu banyak neuron dapat mengakibatkan *overfitting* (Uzun et al., 2024). Berdasarkan beberapa percobaan yang telah dilakukan, berikut

adalah hasil eksplorasi terkait dengan jumlah *hidden layer* dan *nodes* untuk setiap *layer*:

Tabel 11 Hasil eksplorasi jumlah *hidden layer* dan *nodes*

| Jumlah <i>hidden layer</i> | Jumlah <i>nodes</i> | MAE | R ² |
|----------------------------|---------------------|---------|----------------|
| 1 | 128 | 80,8306 | 0,6401 |
| 1 | 256 | 55,0136 | 0,7875 |
| 2 | 128 | 32,6503 | 0,9556 |
| | 128 | | |
| 2 | 256 | 57,9914 | 0,8032 |
| | 256 | | |

Dari percobaan yang telah dilakukan, model mendapatkan hasil paling optimal ketika jumlah *hidden layer* adalah dua dengan jumlah *nodes* masing-masing adalah 128. Jumlah *hidden layer* dan *nodes* ini memberikan waktu komputasi yang wajar dan evaluasi paling baik diantara percobaan lainnya.

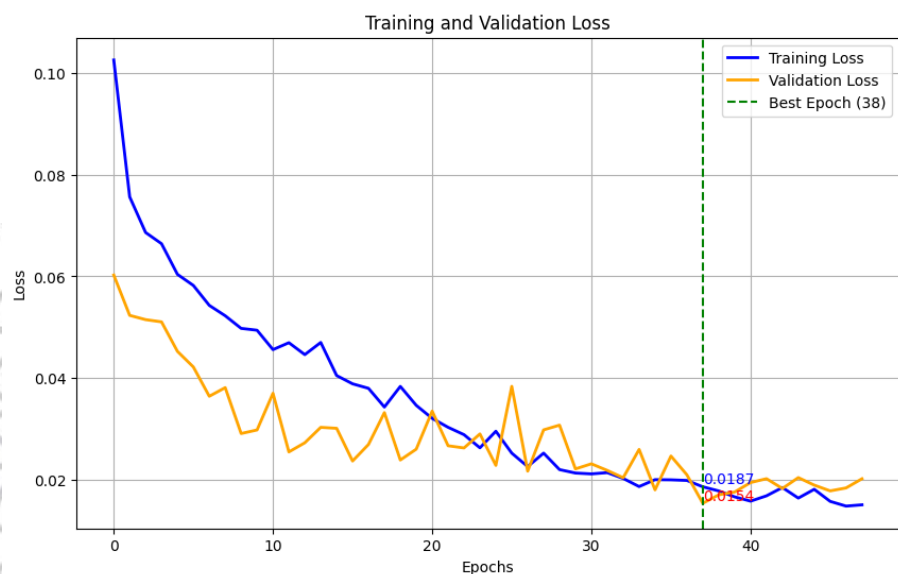
2. Epoch

Jumlah iterasi pelatihan atau *epoch* dalam penelitian ini berjumlah 200 perulangan. Namun, tidak semua perulangan ini akan dilakukan karena adanya penggunaan pustaka *EarlyStopping*. Salah satu parameter yang digunakan dalam pustaka ini adalah *patience*. Parameter ini menjadi acuan bagi *EarlyStopping* untuk menentukan kapan perulangan harus berhenti. Nilai *patience* yang terlalu kecil mungkin dapat menghentikan proses pembelajaran sebelum model konvergen. Namun, nilai *patience* yang terlalu besar akan memakan lebih banyak waktu komputasi dan meningkatkan risiko terjadinya *overfitting*.

Berikut adalah percobaan dalam menemukan nilai optimal untuk parameter *patience*:

Tabel 12 Hasil eksplorasi nilai parameter *patience*

| <i>Patience</i> | <i>Validation Loss</i> | MAE | R^2 |
|-----------------|------------------------|----------------|---------------|
| 5 | 0,0197 | 60,5565 | 0,7614 |
| 10 | 0,0048 | 32,6503 | 0,9556 |
| 15 | 0,0134 | 46,1339 | 0,8376 |



Gambar 19 Grafik *training loss* dan *validation loss*

Berdasarkan hasil eksplorasi nilai parameter *patience* pada Tabel 12 di atas, model mencapai titik optimalnya ketika *patience* berjumlah 10. Sementara Gambar 19 menyatakan bahwa *epoch* terbaik adalah *epoch* ke-38 dengan nilai *training loss* dan *validation loss* paling rendah. Oleh karena itu pada eksplorasi berikutnya nilai *patience* ini akan dipertahankan.

3. Fungsi optimasi

Dalam penelitian yang menggunakan data deret waktu ini, model cenderung bergantung pada pola temporal. Fungsi optimasi RMSProp dan Adam sesuai untuk diuji pada eksplorasi lebih

lanjut karena kemampuan keduanya untuk menyesuaikan bobot secara adaptif. Dilakukan beberapa percobaan dalam menentukan fungsi optimasi yang paling sesuai untuk meningkatkan performa model. Hasil percobaan tersebut dapat diuraikan dalam tabel sebagai berikut:

Tabel 13 Hasil eksplorasi fungsi optimasi

| Fungsi Optimasi | MAE | R ² |
|-----------------|----------------|----------------|
| RMSpop | 53,4797 | 0,8004 |
| Adam | 32,6503 | 0,9556 |

Berdasarkan hasil eksplorasi pada tabel di atas, *optimizer* yang digunakan untuk percobaan-percobaan berikutnya adalah Adam dengan MAE 32,6503 dan R² 0,9556.

4. *Learning rate*

Penelitian ini memanfaatkan pustaka *ReduceLROnPlateau* dimana *learning rate* dapat diatur secara otomatis. Pustaka *ReduceLROnPlateau* bekerja dengan nilai minimum 0,001 dan akan terus mempertahankan nilai tersebut selama *validation loss* masih menunjukkan penurunan. Jika setelah 10 kali perulangan tapi *validation loss* tidak mengalami penurunan maka pustaka *ReduceLROnPlateau* akan mengurangi langkah pembelajaran sebanyak setengah dari nilai saat ini. Hal ini memungkinkan model akan lebih banyak belajar dalam setiap iterasinya.

5. *Timestep*

Model yang dibangun dengan LSTM sangat bergantung kepada nilai *timestep*. *Timestep* yang terlalu pendek mungkin tidak dapat menangkap pola jangka panjang, sedangkan *timestep* yang terlalu panjang dapat meningkatkan kompleksitas komputasi (Furizal et al., 2024).

Adapun beberapa percobaan yang telah dilakukan untuk mendapatkan nilai *timestep* adalah sebagai berikut:

Tabel 14 Hasil eksplorasi nilai *timestep*

| <i>Timestep</i> | MAE | R ² |
|-----------------|----------------|----------------|
| 5 | 81,7322 | 0,6128 |
| 10 | 73,2857 | 0,7075 |
| 20 | 61,8720 | 0,7755 |
| 50 | 32,6503 | 0,9556 |

Dari hasil eksplorasi tersebut maka nilai 50 dipilih sebagai nilai *timestep* yang digunakan pada tahap eksplorasi berikutnya.

6. *Batch size*

Pemilihan nilai *batch size* akan bergantung pada jumlah data. Nilai *batch size* yang lebih besar akan memerlukan lebih banyak memori terutama saat proses latih. Sementara nilai *batch size* yang terlalu kecil akan memakan waktu latih yang lebih lama. Pada penelitian ini dilakukan eksplorasi pada nilai *batch size* yang umum digunakan untuk menemukan nilai yang optimal dan seimbang antara penggunaan memori dan performa model. Hasil percobaan dapat dirincikan dalam tabel berikut ini:

Tabel 15 Hasil eksplorasi nilai *batch size*

| <i>Batch Size</i> | MAE | R ₂ |
|-------------------|----------------|----------------|
| 16 | 56,6038 | 0,7875 |
| 32 | 32,6503 | 0,9556 |
| 64 | 52,8849 | 0,8000 |

Percobaan yang diuraikan dalam Tabel 15 menghasilkan kesimpulan bahwa nilai *batch size* yang menghasilkan performa model paling baik adalah 32. Oleh karenanya, nilai ini akan digunakan dalam percobaan-percobaan berikutnya.

7. Fungsi aktivasi

Pada standarnya, LSTM menggunakan *sigmoid* (σ) sebagai fungsi aktivasi pada *forget gate* dan *tanh* sebagai fungsi aktivasi pada *output gate*. Dalam penelitian ini, dilakukan percobaan terhadap perubahan fungsi aktivasi di *output gate*. ReLU sesuai untuk dieksplorasi karena kemampuannya menangani input negatif. Percobaan dalam eksplorasi fungsi aktivasi dapat diuraikan dalam tabel sebagai berikut:

Tabel 16 Hasil eksplorasi fungsi aktivasi

| Fungsi Aktivasi | MAE | R ² |
|-----------------|----------------|----------------|
| tanh | 69,7545 | 0,7070 |
| ReLU | 32,6503 | 0,9556 |

Berdasarkan Tabel 16, model memiliki performa lebih baik jika menggunakan fungsi aktivasi ReLU. Fungsi aktivasi ini menghasilkan evaluasi MAE 32,6503 dan R² 0,9556.

8. Dropout

Dropout merupakan salah satu teknik yang paling sering digunakan untuk menghindari model *overfitting*. Dalam penelitian ini dilakukan beberapa percobaan dengan nilai *dropout* yang umum digunakan dalam model prediksi sebagaimana dijelaskan dalam tabel berikut ini:

Tabel 17 Hasil eksplorasi nilai *dropout*

| <i>Dropout</i> | MAE | R ² |
|----------------|----------------|----------------|
| 0,2 | 59,6749 | 0,7657 |
| 0,3 | 32,6503 | 0,9556 |
| 0,4 | 62,1817 | 0,7730 |
| 0,5 | 51,7748 | 0,8411 |

Berdasarkan hasil percobaan pada Tabel 17, maka digunakan nilai *dropout* sebesar 0,3 dengan hasil evaluasi MAE 32,6503 dan R² 0,9556.

4.3.2 Arsitektur Model

Model yang digunakan dalam penelitian ini dihasilkan dari uji coba eksperimental. Model dikembangkan agar dapat belajar secara optimal terhadap data yang digunakan. Adapun arsitektur model yang digunakan dapat dilihat pada tabel di bawah ini:

Tabel 18 *Model summary*

| <i>Layer (type)</i> | <i>Output Shape</i> | <i>Parameter</i> |
|---------------------------|---------------------|------------------|
| lstm_8 (LSTM) | (None, 50, 128) | 68,608 |
| dropout_8 (Dropout) | (None, 50, 128) | 0 |
| lstm_9 (LSTM) | (None, 256) | 131,584 |
| dropout_9 (Dropout) | (None, 256) | 0 |
| dense_4 (Dense) | (None, 1) | 129 |
| Total params: 600,965 | | |
| Trainable params: 200,321 | | |
| Non-trainable params: 0 | | |
| Optimizer params: 400,644 | | |

Pembangunan model dimulai dari menentukan jumlah *hidden layer* dan jumlah *node* pada setiap lapisan. Setelah itu, disesuaikan juga beberapa *hyperparameter* untuk mendukung model untuk terus belajar secara optimal. Adapun *hyperparameter* yang digunakan dalam penelitian ini dapat dilihat pada tabel sebagai berikut:

Tabel 19 *Hyperparameter*

| <i>Hyperparameter</i> | <i>Value</i> |
|-----------------------|--|
| <i>Epoch</i> | 200 iterasi dengan <i>EarlyStopping</i> |
| Fungsi optimasi | Adam |
| <i>Learning rate</i> | Mulai dari 0,001 dengan <i>ReduceLROnPlateau</i> |
| <i>Timestep</i> | 50 |
| <i>Batch Size</i> | 32 |
| Fungsi aktivasi | ReLU |
| <i>Dropout</i> | 0,3 |

Hyperparameter pada Tabel 19 didukung oleh pustaka *EarlyStopping* yang bertujuan untuk menghindari *overfitting*. Iterasi *epoch* akan berhenti apabila setelah 10 kali perulangan model tidak mengalami penurunan *validation loss*. Selain itu, penelitian ini juga menggunakan pustaka *ReduceLROnPlateau* untuk mengontrol *learning rate*. Mulanya *learning rate* ini diinisiasi dengan nilai 0,001 lalu jika model tidak mengalami penurunan *validation loss* selama 10 kali iterasi maka *learning rate* akan dikurangi setengahnya.

4.3 Hasil Validasi Menggunakan Group K-Fold Cross Validation

Dalam pembangunan model dengan menggunakan teknik *k-fold cross validation*, nilai *k* harus diinisialisasi sehingga proses latih data akan dibagi sesuai dengan nilai *k* tersebut dengan bagian data uji dan data validasi yang berbeda-beda. Metode *group k-fold* mampu menangani validasi silang pada data yang bersifat kategorikal. Pada penelitian ini, data dapat dikategorikan berdasarkan negara. Dengan *group k-fold cross validation*, data setiap negara berada pada *fold* yang terpisah sehingga nilai *k* dapat diketahui yaitu 14 sesuai dengan jumlah kategori negara. Hasil validasi menggunakan *group k-fold cross validation* dapat dilihat pada Tabel 20 sebagai berikut:

Tabel 20 Hasil *group k-fold cross validation*

| Nilai <i>k</i> | <i>Validation loss</i> | <i>MAE</i> | R^2 |
|-------------------------------|------------------------|----------------|---------------|
| Tanpa <i>cross validation</i> | 0,0145 | 284,88 | 0,5325 |
| <i>k</i> = 14 | 0,0048 | 32,6503 | 0,9556 |

Dari Tabel 20 di atas, dapat diketahui bahwa teknik *group k-fold cross validation* mempengaruhi kinerja model secara signifikan. Model dengan *group k-fold cross validation* memiliki *validation loss* 0,0048 dengan MAE 32,6503 dan nilai R^2 0,9556.

4.4 Hasil Evaluasi Model

4.4.1 Mean Absolute Error (MAE)

Model akhir dengan penyesuaian arsitektur dan *hyperparameter* seperti pada Tabel 18 dan Tabel 19 mencapai evaluasi pada nilai MAE sebesar 32,6503 yang menunjukkan perbedaan atau selisih antara nilai aktual dan nilai yang diprediksi model. Berdasarkan penelitian yang dilakukan oleh Xiao et al., nilai MAE yang lebih kecil membuktikan performa model yang lebih baik (Xiao et al., 2021). Apabila meninjau kembali dari distribusi data target yang memiliki nilai minimum 8,5 dan maksimum 5377,4 nilai MAE 32,6503 relatif kecil terhadap distribusi data.

4.4.2 R-squared (R^2)

Adapun nilai evaluasi R^2 pada model akhir mencapai 0,9556 yang menunjukkan bahwa model memiliki kinerja yang cukup baik dalam menjelaskan variabilitas dalam data mengingat pada penelitian Plevris et al., disebutkan bahwa kualitas model prediksi semakin baik jika nilai R^2 mendekati satu (Plevris et al., 2022). Dengan demikian, hasil evaluasi ini memberikan indikasi yang jelas tentang kekuatan hubungan antara variabel yang dianalisis, serta efektivitas model dalam memprediksi nilai yang diinginkan.

4.5 Hasil Prediksi Konsumsi Energi Baru Terbarukan

Setelah menerapkan arsitektur dan *hyperparameter* yang didapatkan dari hasil eksplorasi, hasil prediksi yang diperoleh model menunjukkan kinerja yang cukup baik. Sebagai perbandingan, telah dilakukan juga uji coba untuk menerapkan arsitektur dan *hyperparameter* yang sama. Detail perbandingan dapat dilihat pada tabel sebagai berikut:

Tabel 21 Hasil prediksi konsumsi energi baru terbarukan

| Algoritma | Running time | Validation loss | MAE | R^2 |
|---------------|--------------------------|-----------------|----------------|---------------|
| Bidirectional | 30 menit 8 detik | 0,0056 | 33,1371 | 0,9247 |
| LSTM | 18 menit 31 detik | 0,0048 | 32,6503 | 0,9556 |
| RNN | 10 menit 14 detik | 0,0103 | 55,0471 | 0,8328 |

Berdasarkan Tabel 21, dapat disimpulkan bahwa LSTM melakukan prediksi konsumsi energi baru terbarukan dengan performa model yang efektif dan waktu komputasi yang efisien. Berikut merupakan sampel data aktual dan data prediksi yang dihasilkan model:

Tabel 22 Tabel perbandingan nilai aktual dan nilai prediksi

| Indeks | Nilai Aktual | Nilai Prediksi |
|--------|--------------|----------------|
| 0 | 198,0068 | 192,4811 |
| 1 | 198,0068 | 195,8276 |
| 2 | 198,0068 | 210,9728 |
| 3 | 237,8008 | 226,0890 |
| 4 | 359,6926 | 402,2522 |

Berdasarkan Tabel 22, indeks data ke-4 menunjukkan bahwa terdapat perbedaan yang cukup jauh antara nilai aktual dan prediksi. Hal ini merupakan perbedaan yang ditunjukkan oleh hasil evaluasi MAE. Namun demikian, sampel data lainnya memberikan gambaran bahwa model dapat memprediksi konsumsi energi baru terbarukan dengan selisih yang wajar dengan nilai aktualnya.

4.6 Hasil Pengujian Model Prediksi

Berdasarkan model prediksi yang telah dirancang dan diuji melalui berbagai tahap validasi serta evaluasi, langkah berikutnya adalah menguji performa model dengan menggunakan sejumlah sampel input. Hasil prediksi model kemudian akan dibandingkan dengan data aktual untuk mengevaluasi sejauh mana model mampu merepresentasikan pola data yang sebenarnya. Proses ini bertujuan untuk memastikan bahwa model tidak hanya akurat selama pelatihan tetapi juga dapat diaplikasikan secara efektif pada data baru yang belum pernah dilihat sebelumnya.

Dengan digunakannya teknik *group k-fold cross validation* sebesar 14 *fold* model dengan evaluasi terbaik diperoleh dari *fold* ke-7, maka dapat diketahui data latih dan data uji yang digunakan adalah sebagai berikut:

$$X.shape = (700, 5)$$

$$X_lstm_fold = \frac{700}{14} = 50 \text{ baris data untuk setiap fold}$$

Maka, data uji yang digunakan adalah:

$$X \text{ val indeks ke } - 0 = (50 \times 6) + 1 = 301$$

$$X \text{ val indeks ke } - 49 = 301 + 49 = 350$$

Artinya data uji pada *fold* ke-3 berasal dari data baris ke 301 hingga 350 dimana diantara baris tersebut terdapat data tahunan negara Myanmar yang dapat menjadi sampel data uji. Hasil pengujian model prediksi ditampilkan pada Tabel 23 sebagai berikut:

Tabel 23 Hasil pengujian model prediksi

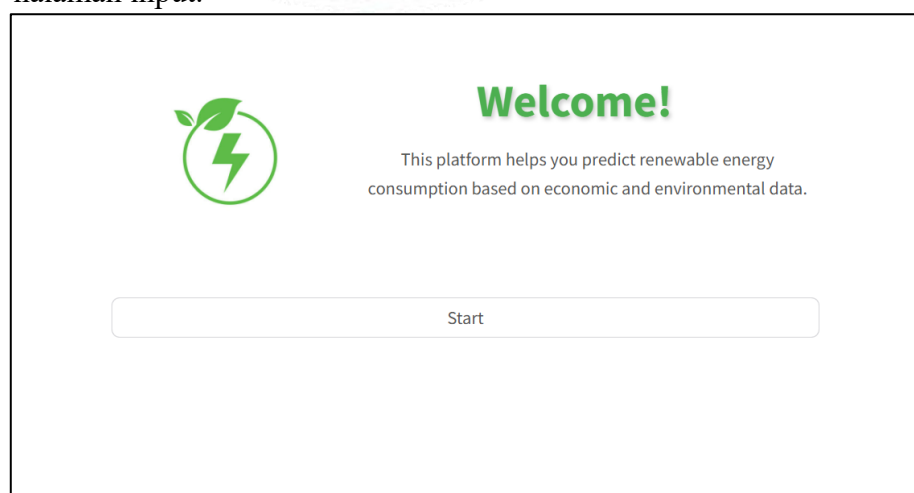
| Country | Year | GDP | FDI_in | CO2_emission | FF_cons | Energy use | Aktual 2023 | Prediksi 2023 |
|---------|------|---------|---------------|--------------|---------|------------|-------------|---------------|
| Myanmar | 2022 | 1149,02 | 1238500000,00 | 0,65 | 3065.85 | 3,064.85 | 175,73 | 207,68 |

Berdasarkan hasil pengujian model prediksi menggunakan data uji, hasil prediksi memiliki selisih yang cukup jauh terhadap nilai aktual. Hal ini menjelaskan alasan hasil evaluasi MAE yang cukup tinggi.

4.7 Hasil Integrasi Sistem

4.7.1 Halaman awal

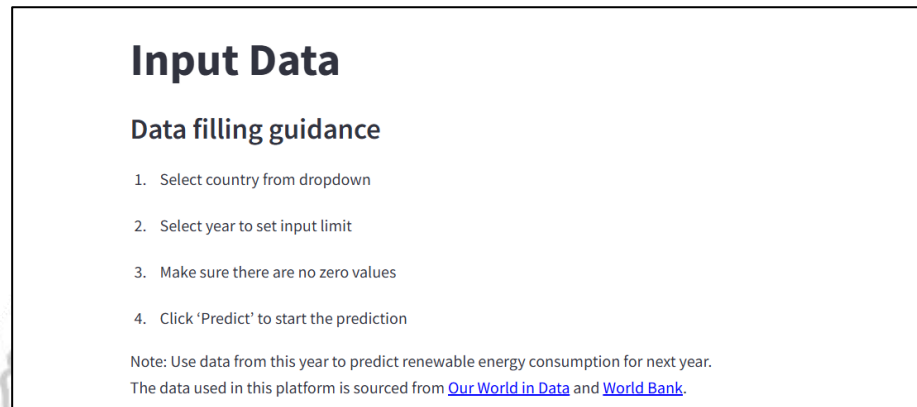
Halaman awal terdiri dari beberapa teks yang menjelaskan isi dari web, logo dan juga tombol untuk memulai prediksi. Ketika pengguna menekan tombol “Start” maka web akan melanjutkan proses ke halaman input.



Gambar 20 Halaman awal web

4.7.2 Halaman input

Pada halaman input, pengguna dapat melihat panduan untuk mengisi data lalu mengikuti instruksinya. Hasil prediksi konsumsi energi baru terbarukan adalah nilai prediksi untuk satu tahun setelah tahun yang dijadikan input oleh pengguna. Selain itu, kolom input memiliki ketentuan tersendiri yaitu nilai tidak boleh sama dengan nol. Hal ini disampaikan dalam bagian panduan pengisian data atau seperti yang tertera pada Gambar 21



Input Data

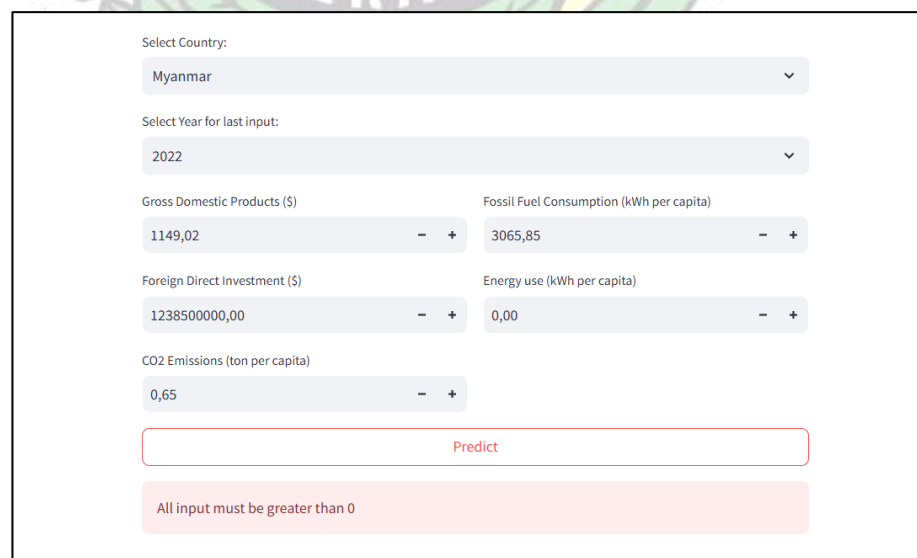
Data filling guidance

1. Select country from dropdown
2. Select year to set input limit
3. Make sure there are no zero values
4. Click 'Predict' to start the prediction

Note: Use data from this year to predict renewable energy consumption for next year.
The data used in this platform is sourced from [Our World in Data](#) and [World Bank](#).

Gambar 21 Panduan pengisian data

Jika pengguna tetap mengisi kolom input dengan nilai nol maka akan muncul pesan kesalahan seperti pada Gambar 22.



Select Country:
Myanmar

Select Year for last input:
2022

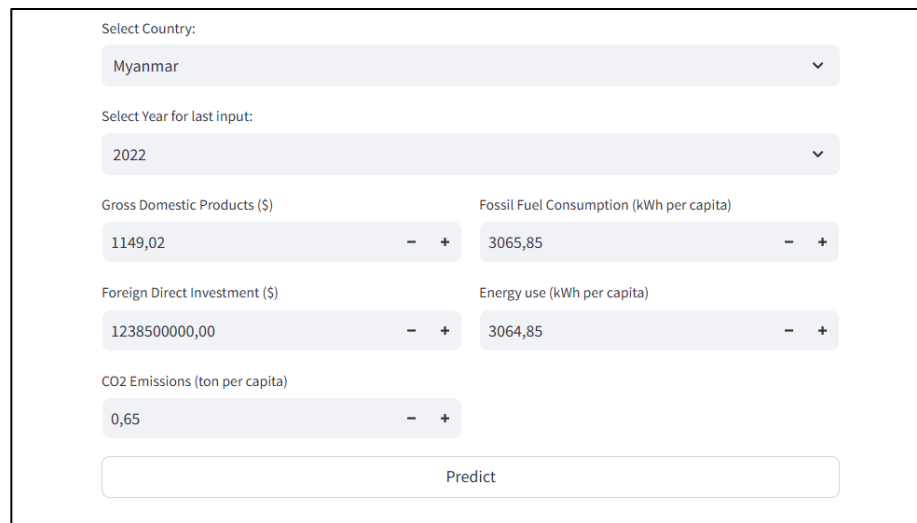
| | | | | | | | |
|--------------------------------|---------------|---|---|--|---------|---|---|
| Gross Domestic Products (\$) | 1149,02 | - | + | Fossil Fuel Consumption (kWh per capita) | 3065,85 | - | + |
| Foreign Direct Investment (\$) | 1238500000,00 | - | + | Energy use (kWh per capita) | 0,00 | - | + |
| CO2 Emissions (ton per capita) | 0,65 | - | + | | | | |

Predict

All input must be greater than 0

Gambar 22 Pesan kesalahan pada halaman input

Setelah semua kolom input diisi, maka selanjutnya pengguna dapat memilih tombol “*Predict*” untuk melihat hasil prediksi.



Select Country:
Myanmar

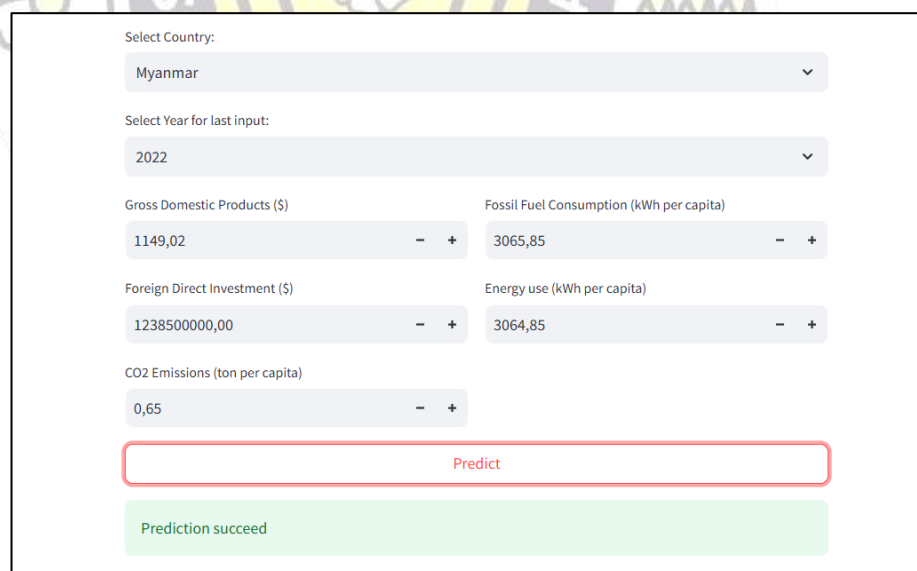
Select Year for last input:
2022

| | | | | | | | |
|--------------------------------|---------------|---|---|--|---------|---|---|
| Gross Domestic Products (\$) | 1149,02 | - | + | Fossil Fuel Consumption (kWh per capita) | 3065,85 | - | + |
| Foreign Direct Investment (\$) | 1238500000,00 | - | + | Energy use (kWh per capita) | 3064,85 | - | + |
| CO2 Emissions (ton per capita) | 0,65 | - | + | | | | |

Predict

Gambar 23 Halaman input web

Jika prediksi berhasil dilakukan, maka akan muncul pesan bahwa prediksi berhasil dilakukan seperti pada Gambar 24 berikut:



Select Country:
Myanmar

Select Year for last input:
2022

| | | | | | | | |
|--------------------------------|---------------|---|---|--|---------|---|---|
| Gross Domestic Products (\$) | 1149,02 | - | + | Fossil Fuel Consumption (kWh per capita) | 3065,85 | - | + |
| Foreign Direct Investment (\$) | 1238500000,00 | - | + | Energy use (kWh per capita) | 3064,85 | - | + |
| CO2 Emissions (ton per capita) | 0,65 | - | + | | | | |

Predict

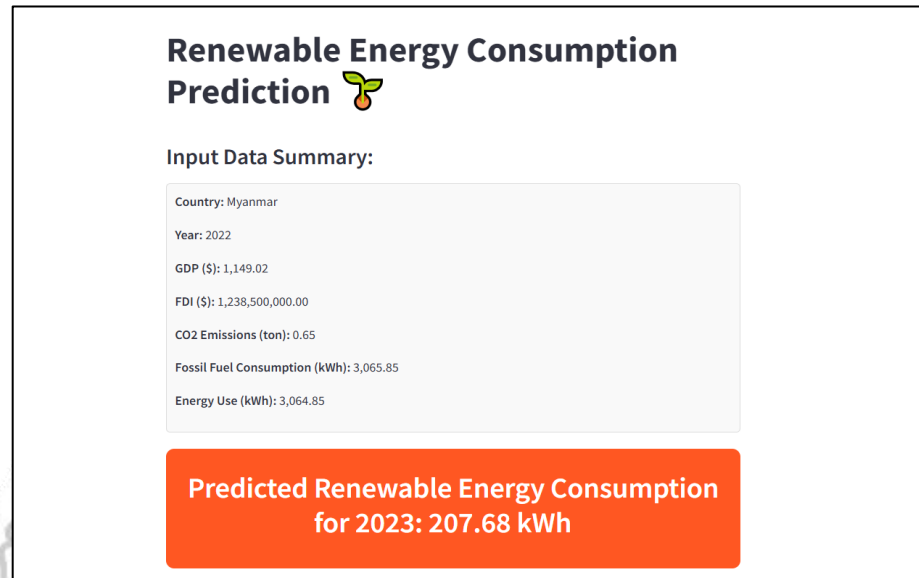
Prediction succeed

Gambar 24 Pesan prediksi berhasil

Selanjutnya, pengguna akan diarahkan pada halaman output yang menampilkan hasil prediksi konsumsi energi baru terbarukan.

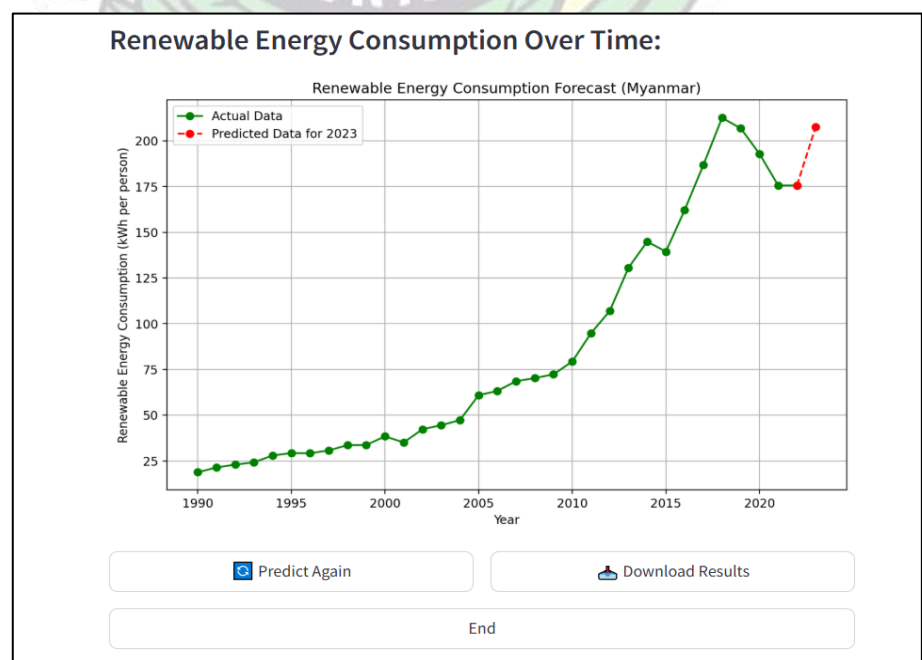
4.7.3 Halaman output

Pada halaman output, ditampilkan ringkasan data input untuk memastikan pengguna dapat melihat kembali bahwa data inputnya telah diisi dengan benar. Hasil prediksi disorot dengan warna terang untuk meningkatkan keterbacaan oleh pengguna.



Gambar 25 Halaman output web

Selain hasil prediksi konsumsi energi baru terbarukan dalam satuan kWh, pengguna juga dapat melihat tren konsumsi energi baru terbarukan dari tahun ke tahun. Pengguna dapat menemukan grafik seperti pada gambar berikut:

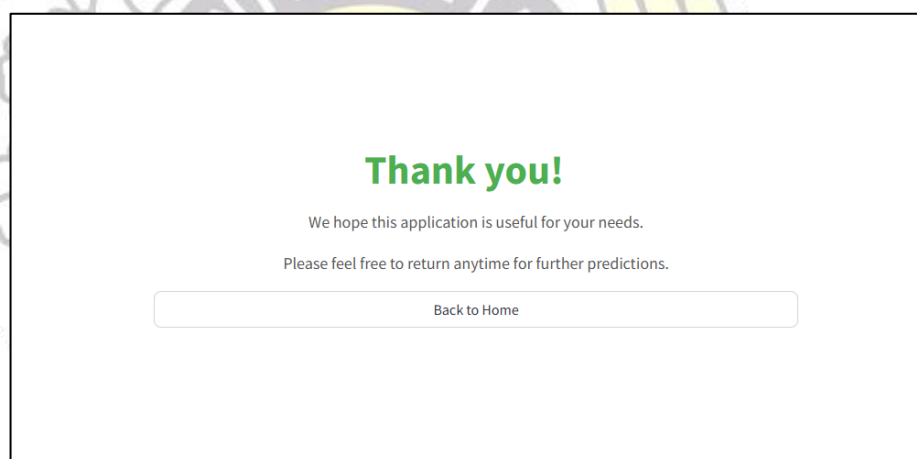


Gambar 26 Grafik tren konsumsi energi baru terbarukan

Garis hijau pada grafik merujuk pada data aktual yang digunakan untuk melatih model prediksi. Sedangkan garis putus-putus dan titik merah merupakan hasil prediksi konsumsi energi baru terbarukan yang telah dihitung oleh sistem berdasarkan input pengguna. Pengguna juga dapat memilih tombol *“Predict Again”* untuk mengembalikan halaman input. Tombol *“Download Result”* menghasilkan dokumen dalam ekstensi .txt yang mengandung informasi nilai input pengguna dan hasil prediksinya. Lalu tombol *“End”* akan membawa pengguna menuju halaman akhir.

4.7.4 Halaman akhir

Halaman akhir berisi teks yang mengakhiri sistem dalam melakukan sistem prediksi. Jika tombol *“Back to Home”* ditekan maka ini akan mengembalikan sistem pada halaman awal.



Gambar 27 Halaman akhir web

BAB 5 PENUTUP

5.1 Kesimpulan

Adapun kesimpulan yang dapat diambil dari penelitian ini yaitu:

1. Algoritma *Long Short Term Memory (LSTM)* telah berhasil diimplementasikan untuk membangun model prediksi konsumsi energi baru terbarukan dengan data tahunan ekonomi dan lingkungan dari 14 negara berkembang di Asia termasuk diantaranya Bangladesh, Brunei Darussalam, China, Filipina, India, Indonesia, Iran, Kamboja, Laos, Malaysia, Myanmar, Pakistan, Thailand dan Vietnam. Model dilatih untuk dapat memprediksi angka konsumsi energi baru terbarukan untuk tahun depan berdasarkan data tahun sebelumnya.
2. Hasil evaluasi model diinterpretasikan dengan metrik evaluasi *Mean Absolute Error (MAE)* sebesar 32,6503 dan R^2 sebesar 0,9556. Arsitektur model yang digunakan adalah sebagai berikut:
 - a. Jumlah *hidden layer*: 2
 - b. Jumlah neuron *hidden layer* pertama: 128
 - c. Jumlah neuron *hidden layer* kedua: 128
 - d. *Epoch*: 200 dengan *EarlyStopping*
 - e. Fungsi optimasi: Adam
 - f. *Timestep*: 50
 - g. *Batch size*: 32
 - h. Fungsi aktivasi: ReLU
 - i. *Dropout*: 0.3
 - j. *Group k-fold cross validation*: 14 fold

5.2 Saran

Berdasarkan hasil penelitian, berikut adalah beberapa saran yang dapat dipertimbangkan untuk penelitian selanjutnya:

1. Berdasarkan penelitian yang telah dilakukan, model terbilang cukup baik untuk tahap awal. Namun penelitian ini masih memiliki ruang untuk pengembangan lebih lanjut. Peneliti menyarankan agar penelitian selanjutnya dapat melakukan pengumpulan data yang lebih mendalam sehingga model dapat memperoleh akurasi yang lebih baik.
2. Pada penelitian selanjutnya, tahap eksplorasi *hyperparameter* dapat dilakukan dengan metode lain seperti *Bayesian Optimization*, *Tree-structured Parzen Estimator (TPE)* atau pendekatan lainnya yang sesuai digunakan pada data *time series*.

DAFTAR PUSTAKA

- Abhigyan. (2021). *Cross-Validation Techniques*. Medium.
<https://medium.com/geekculture/cross-validation-techniques-33d389897878>
- Adi, A. C. (2024). *Kementerian ESDM Gaet Investasi, Tingkatkan Bauran Energi dari EBT*. Kementrian ESDM. <https://www.esdm.go.id/id/media-center/arsip-berita/kementerian-esdm-gaet-investasi-tingkatkan-bauran-energi-dari-ebt>
- Ang, T. Z., Salem, M., Kamarol, M., Das, H. S., Nazari, M. A., & Prabaharan, N. (2022). A comprehensive study of renewable energy sources: Classifications, challenges and suggestions. *Energy Strategy Reviews*, 43(August), 100939. <https://doi.org/10.1016/j.esr.2022.100939>
- Bérchez-Moreno, F., Fernández, J. C., Hervás-Martínez, C., & Gutiérrez, P. A. (2024). Fusion of standard and ordinal dropout techniques to regularise deep models. *Studies in Big Data*, 46(October 2023), 61–78. https://doi.org/10.1007/978-3-319-97556-6_4
- Chávez, P., & Martini, I. (2024). Applying neural networks for short and long-term hourly electricity consumption forecasting in universities: A simultaneous approach for energy management. *Journal of Building Engineering*, 97(June). <https://doi.org/10.1016/j.jobe.2024.110612>
- Dash, C. S. K., Behera, A. K., Dehuri, S., & Ghosh, A. (2023). An outliers detection and elimination framework in classification task of data mining. *Decision Analytics Journal*, 6(January). <https://doi.org/10.1016/j.dajour.2023.100164>
- Dolgintseva, E., Wu, H., Petrosian, O., Zhadan, A., Allakhverdyan, A., & Martemyanov, A. (2024). Comparison of multi-step forecasting methods for renewable energy. In *Energy Systems* (Issue 0123456789). <https://doi.org/10.1007/s12667-024-00656-w>
- Dubey, S. R., Singh, S. K., & Chaudhuri, B. B. (2022). Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing*, 503, 92–108. <https://doi.org/10.1016/j.neucom.2022.06.111>
- Fikriyyah, A. K., & Boedoyo, M. S. (2021). Analysis of renewable energy implementation in remote areas of Indonesia: Study review. *IOP Conference Series: Earth and Environmental Science*, 673(1). <https://doi.org/10.1088/1755-1315/673/1/012003>

- Flinn, G. (2024). *What if we ran out of fossil fuels?* HowStuffWorks. <https://science.howstuffworks.com/science-vs-myth/what-if/what-if-ran-out-fossil-fuels.htm>
- Furizal, Ma'arif, A., Suwarno, I., Masitha, A., Aulia, L., & Sharkawy, A. N. (2024). Real-Time Mechanism Based on Deep Learning Approaches for Analyzing the Impact of Future Timestep Forecasts on Actual Air Quality Index of PM10. *Results in Engineering*, 24(November). <https://doi.org/10.1016/j.rineng.2024.103434>
- He, Y., Zhao, J., Yao, L., & Li, S. (2024). A deep learning-based ground motion truncation method to improve efficiency of structural time history analysis. *Structures*, 63(April). <https://doi.org/10.1016/j.istruc.2024.106381>
- Heidari, M., Moattar, M. H., & Ghaffari, H. (2023). Forward propagation dropout in deep neural networks using Jensen–Shannon and random forest feature importance ranking. *Neural Networks*, 165, 238–247. <https://doi.org/10.1016/j.neunet.2023.05.044>
- Hoa, P. X., Xuan, V. N., & Phuong Thu, N. T. (2023). Determinants of the renewable energy consumption: The case of Asian countries. *Heliyon*, 9(12), e22696. <https://doi.org/10.1016/j.heliyon.2023.e22696>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Huang, H., Huang, X., Ding, W., Zhang, S., & Pang, J. (2023). Optimization of electric vehicle sound package based on LSTM with an adaptive learning rate forest and multiple-level multiple-object method. *Mechanical Systems and Signal Processing*, 187(November 2022). <https://doi.org/10.1016/j.ymssp.2022.109932>
- Kementrian Energi dan Sumber Daya Mineral. (2023). *Handbook Of Energy & Economic Statistics Of Indonesia*. <https://esdm.go.id/id/publikasi/handbook-of-energy-economic-statistics-of-indonesia>
- Kocatas, U. (2024). *Missing Data Imputation Aproaches and Application*. Medium. <https://medium.com/@umut.kocatas41/missing-data-imputation-aproaches-and-application-e3f9d1b06f98>
- Kristiyanti, D. A., Pramudya, W. B. N., & Sanjaya, S. A. (2024). How can we predict transportation stock prices using artificial intelligence? Findings from experiments with Long Short-Term Memory based algorithms. *International Journal of Information Management Data Insights*, 4(2). <https://doi.org/10.1016/j.jjime.2024.100293>

- Leni, D. (2023). The Influence of Heatmap Correlation-based Feature Selection on Predictive Modeling of Low Alloy Steel Mechanical Properties Using Artificial Neural Network (ANN) Algorithm. *Journal of Energy, Material, and Instrumentation Technology*, 4(4), 152–162. <https://doi.org/10.23960/jemit.v4i4.203>
- Meisenbacher, S., Turowski, M., Phipps, K., Rätz, M., Müller, D., Hagenmeyer, V., & Mikut, R. (2022). Review of automated time series forecasting pipelines. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(6). <https://doi.org/10.1002/widm.1475>
- Meng, Z., Sun, H., & Wang, X. (2022). Forecasting Energy Consumption Based on SVR and Markov Model: A Case Study of China. *Frontiers in Environmental Science*, 10(April), 1–15. <https://doi.org/10.3389/fenvs.2022.883711>
- Minister of Energy and Mineral Resources of the Republic of Indonesia. (2023). *General Plan for National Electricity. 1*, 1–271.
- Muta, T., & Erdogan, M. (2023). *The global energy crisis pushed fossil fuel consumption subsidies to an all-time high*. IEA. <https://www.iea.org/commentaries/the-global-energy-crisis-pushed-fossil-fuel-consumption-subsidies-to-an-all-time-high>
- Nonthapot, S., Lean, H. H., Watchalaanun, T., & Sihabutr, C. (2024). The mediating effect of economic factors and energy consumption between tourism and carbon emissions in ASEAN countries. *Journal of Open Innovation: Technology, Market, and Complexity*, 10(2). <https://doi.org/10.1016/j.joitmc.2024.100270>
- Obatola, S. O., & Junjie, T. (2024). A data-driven approach to grid-connected PV system reliability assessment: Combining deep learning and hybrid optimization. *Energy Reports*, 12(November), 5582–5593. <https://doi.org/10.1016/j.egyr.2024.11.041>
- Pambudi, N. A., Firdaus, R. A., Rizkiana, R., Ulfa, D. K., Salsabila, M. S., Suharno, & Sukatiman. (2023). Renewable Energy in Indonesia: Current Status, Potential, and Future Development. *Sustainability (Switzerland)*, 15(3). <https://doi.org/10.3390/su15032342>
- Plevris, V., Solorzano, G., Bakas, N. P., & Ben Seghier, M. E. A. (2022). Investigation of Performance Metrics in Regression Analysis and Machine Learning-Based Prediction Models. *World Congress in Computational Mechanics and ECCOMAS Congress, June*. <https://doi.org/10.23967/eccomas.2022.155>

- Pratiwi, S., & Juerges, N. (2020). Review of the impact of renewable energy development on the environment and nature conservation in Southeast Asia. *Energy, Ecology and Environment*, 5(4), 221–239. <https://doi.org/10.1007/s40974-020-00166-2>
- Ritchie, H., & Rosado, P. (2024). *Fossil Fuels*. Our World in Data. <https://ourworldindata.org/fossil-fuels>
- Talwariya, A., Singh, P., Jobanputra, J. H., & Kolhe, M. L. (2023). Machine learning based renewable energy generation and energy consumption forecasting. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*, 45(2), 3266–3278. <https://doi.org/10.1080/15567036.2023.2194257>
- Udin, U. (2020). Renewable energy and human resource development: Challenges and opportunities in Indonesia. *International Journal of Energy Economics and Policy*, 10(2), 233–237. <https://doi.org/10.32479/ijeep.8782>
- Uzun, S., Yildiz, E., & Arslantaş, H. (2024). Optimizing neural network models for predicting nuclear reactor channel temperature: A study on hyperparameter tuning and performance analysis. *Nuclear Engineering and Design*, 429(September). <https://doi.org/10.1016/j.nucengdes.2024.113636>
- Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review on the long short-term memory model. *Artificial Intelligence Review*, 53(8), 5929–5955. <https://doi.org/10.1007/s10462-020-09838-1>
- Xiao, Q., Li, C., Member, S., Tang, Y., & Member, S. (2021). *Energy Efficiency Modeling for Configuration-Dependent Machining via Machine Learning: A Comparative Study*. 18(2), 717–730.