

**PREDIKSI PENGGUNAAN AIR BERSIH PERUSAHAAN
UMUM DAERAH AIR MINUM MENGGUNAKAN
ALGORITMA *LONG SHORT TERM MEMORY***

LAPORAN TUGAS AKHIR

Laporan ini disusun guna memenuhi salah satu syarat untuk menyelesaikan
program studi S1 Teknik Informatika pada Fakultas Teknologi Industri Universitas
Islam Sultan Agung Semarang



DISUSUN OLEH :

HENDRI KURNIAWAN

32602100050

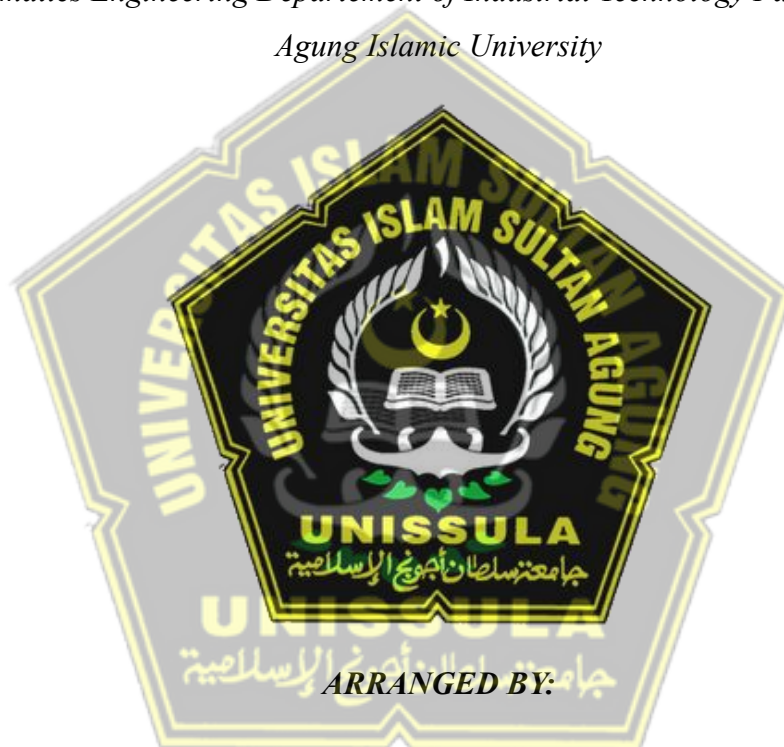
**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM SULTAN AGUNG
SEMARANG**

2025

***PREDICTION OF CLEAN WATER USE BY REGIONAL PUBLIC
DRINKING WATER COMPANIES USING LONG SHORT TERM
MEMORY ALGORITHM***

FINAL PROJECT

*Proposed to complete the requirement to obtain a bachelor's degree (S1) at
Informatics Engineering Departement of Industrial Technology Faculty Sultan
Agung Islamic University*



ARRANGED BY:

HENDRI KURNIAWAN

32602100050

***MAJORING OF INFORMATICS ENGINEERING
INDUSTRIAL TECHNOLOGY FACULTY
SULTAN AGUNG ISLAMIC UNIVERSITY
SEMARANG
2025***

LEMBAR PENGESAHAN TUGAS AKHIR

**PREDIKSI PENGGUNAAN AIR BERSIH PERUSAHAAN UMUM
DAERAH AIR MINUM MENGGUNAKAN ALGORITMA *LONG SHORT
TERM MEMORY***

HENDRI KURNIAWAN

NIM 32602100050

Telah dipertahankan di depan tim penguji ujian sarjana tugas akhir

Program Studi Teknik Informatika

Universitas Islam Sultan Agung

Pada tanggal : 01 September 2025

TIM PENGUJI UJIAN SARJANA :

Moch Taufik, ST.,MIT.

NIK. 210604034

(Penguji 1)

Imam M. Ibnu Subroto, ST., MSC.,

PH.D

NIK. 210600017

(Penguji 2)

Bagus SWP, S.Kom, M.Cs

NIK. 1027118801

(Pembimbing)

02-09-2025

02-09-2025

02-09-2025

Semarang, 2 September 2025

Mengetahui,

Kaprodi Teknik Informatika

Universitas Islam Sultan Agung

Moch. Taufik, ST, MIT

NIDN. 0622037502

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Hendri Kurniawan

NIM : 32602100050

Judul Tugas Akhir : PREDIKSI PENGGUNAAN AIR BERSIH PERUSAHAAN
UMUM DAERAH AIR MINUM MENGGUNAKAN ALGORITMA *LONG
SHORT TERM MEMORY*

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 2 September 2025

Yang Menyatakan,



Hendri Kurniawan

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Hendri Kurniawan

NIM : 32602100050

Program Studi : Teknik Informatika

Fakultas : Teknologi industri

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul : Prediksi Penggunaan Air Bersih Perusahaan Umum Daerah Air Minum Menggunakan Algoritma *Long Short Term Memory*

Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan pangkalan data dan dipublikasikan diinternet dan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan agung.

Semarang, 2 September 2025

Yang menyatakan,



Hendri Kurniawan

KATA PENGANTAR

Dengan mengucapkan syukur alhamdulillah atas kehadiran Allah SWT yang telah memberikan rahmat dan karunianya kepada penulis, sehingga dapat menyelesaikan Tugas Akhir dengan judul “Prediksi Penggunaan Air Bersih Perusahaan Umum Daerah Air Minum Menggunakan Algoritma *Long Short Term Memory*” ini untuk memenuhi salah satu syarat menyelesaikan studi serta dalam rangka memperoleh gelar sarjana (S-1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang.

Tugas Akhir ini disusun dan dibuat dengan adanya bantuan dari berbagai pihak, materi maupun teknis, oleh karena itu saya selaku penulis mengucapkan terima kasih kepada:

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, S.H., M.H yang mengizinkan penulis menimba ilmu di kampus ini.
2. Dekan Fakultas Teknologi Industri Ibu Dr. Novi Marlyana, S.T., M.T.
3. Dosen pembimbing I penulis Bapak Bagus SWP, S.Kom, M.Cs yang telah meluangkan waktu dan memberi ilmu. Serta memberikan banyak nasehat dan saran.
4. Orang tua penulis yang telah mengizinkan untuk menyelesaikan laporan ini,
5. Dan kepada semua pihak yang tidak dapat saya sebutkan satu persatu.

Dengan segala kerendahan hati, penulis menyadari masih terdapat banyak kekurangan dari segi kualitas atau kuantitas maupun dari ilmu pengetahuan dalam penyusunan laporan, sehingga penulis mengharapkan adanya saran dan kritikan yang bersifat membangun demi kesempurnaan laporan ini dan masa mendatang.

Semarang, 2 September 2025



Hendri Kurniawan

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN TUGAS AKHIR	iii
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR.....	iv
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
ABSTRAK	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Tujuan.....	3
1.5 Manfaat	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	5
2.1 Tinjauan Pustaka	5
2.2 Dasar Teori	7
2.2.1 Air Bersih PDAM.....	7
2.2.2 Sistem Prediksi	9
2.2.3 <i>Machine Learning</i>	11
2.2.4 <i>Long Short Term Memory</i>	12
BAB III METODE PENELITIAN	16
3.1 Metode Penelitian.....	16
3.1.1 Studi Literatur.....	16
3.1.2 Pengumpulan Data.....	17
3.1.3 <i>Preprocessing Data</i>	17
3.1.4 Model LSTM	18
3.1.5 Evaluasi Model	21

3.1.6	Visualisasi Hasil.....	23
3.1.7	Rancangan GUI.....	23
3.2	Software Yang Digunakan.....	24
BAB IV HASIL DAN ANALISIS PENELITIAN.....		31
4.1	Hasil Pengumpulan Data.....	31
4.2	Hasil <i>Preprocessing</i> Data.....	32
4.3	Implementasi Pelatihan Model.....	34
4.4	Evaluasi Model.....	35
4.5	Visualisasi Hasil.....	40
4.6	Hasil Implementasi.....	43
BAB V KESIMPULAN DAN SARAN		49
5.1	Kesimpulan	49
5.2	Saran.....	49
DAFTAR PUSTAKA		



DAFTAR GAMBAR

Gambar 2. 1 arsitektur LSTM	13
Gambar 3. 1 <i>Flowchart</i> Penelitian	16
Gambar 3. 2 Dataset Penggunaan Air	17
Gambar 3. 3 Flowchart Perancangan Model.....	18
Gambar 3. 4 Rancangan GUI	23
Gambar 4. 1 Hasil Visualisasi Wilayah 1	41
Gambar 4. 2 Visualisasi Hasil Wilayah 2	42
Gambar 4. 3 Tampilan awal sistem	44
Gambar 4. 4 Opsi Prediksi	45
Gambar 4. 5 Hasil Prediksi	45
Gambar 4. 6 Tabel Prediksi	46
Gambar 4. 7 Visualisasi grafik	47
Gambar 4. 8 Pilihan Wilayah	47



DAFTAR TABEL

Tabel 4. 1 Hasil pengumpulan data	31
Tabel 4. 2 Data hilang	32
Tabel 4. 3 Data duplikat	33
Tabel 4. 4 Hasil Evaluasi Pemakaian Air	36
Tabel 4. 5 Hasil Evaluasi Jumlah Pelanggan	38



ABSTRAK

Perusahaan Umum Daerah Air Minum (PDAM) berperan penting dalam penyediaan air bersih bagi masyarakat, sehingga diperlukan prediksi kebutuhan air yang akurat untuk mendukung distribusi yang optimal dan berkelanjutan. Penelitian ini bertujuan membangun model prediksi penggunaan air bersih dengan algoritma Long Short-Term Memory (LSTM), salah satu jenis Recurrent Neural Network (RNN) yang efektif dalam menganalisis data deret waktu. Data yang digunakan berupa catatan historis jumlah pelanggan dan volume pemakaian air bersih dari 15 wilayah layanan PDAM Kabupaten Kendal periode 2020–2025. Tahapan penelitian mencakup pra-pemrosesan data, perancangan dan pelatihan model, evaluasi performa, serta implementasi ke dalam sistem berbasis web menggunakan Streamlit. Hasil evaluasi menunjukkan bahwa model LSTM mampu menghasilkan prediksi dengan tingkat akurasi tinggi, ditunjukkan oleh nilai Mean Absolute Percentage Error (MAPE) rata-rata 3–5% pada pemakaian air dan kurang dari 2% pada jumlah pelanggan, serta nilai Root Mean Squared Error (RMSE) yang relatif rendah. Implementasi berbasis web memungkinkan PDAM melakukan analisis prediksi secara interaktif dan mendukung perencanaan distribusi air yang lebih efisien.

Kata Kunci: LSTM, prediksi, PDAM, penggunaan air bersih, deret waktu.

ABSTRACT

The Regional Drinking Water Company (PDAM) plays an essential role in providing clean water services for the community, making accurate demand forecasting crucial to ensure optimal and sustainable distribution. This study aims to develop a clean water consumption prediction model using the Long Short-Term Memory (LSTM) algorithm, a variant of Recurrent Neural Network (RNN) effective in analyzing time series data. The dataset consists of historical records of customer numbers and water consumption volumes from 15 service areas of PDAM Kendal Regency during the 2020–2025 period. The research stages include data preprocessing, model design and training, performance evaluation, and system implementation through a web-based application using Streamlit. The evaluation results show that the LSTM model provides high prediction accuracy, with an average Mean Absolute Percentage Error (MAPE) of 3–5% for water consumption and less than 2% for customer numbers, along with relatively low Root Mean Squared Error (RMSE) values. The web-based implementation enables PDAM to perform interactive prediction analysis and supports more efficient water distribution planning.

Keywords: LSTM, prediction, PDAM, clean water consumption, time series.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Air termasuk kebutuhan pokok yang sangat esensial bagi kelangsungan hidup manusia dan makhluk hidup lainnya. Air selama ini sering dipandang sebagai sumber daya alam yang tidak akan habis dan selalu bisa dimanfaatkan kapan pun diperlukan. Namun, pada kenyataannya air tidak selalu tersedia dalam jumlah cukup, karena pergerakannya mengikuti siklus alam yang cenderung tetap, pasokan air jadi terbatas dan sulit menyesuaikan dengan kebutuhan yang terus meningkat (Djana, 2023). Air sangat penting dalam kebutuhan pokok yang diperlukan kehidupan makhluk hidup salah satunya manusia untuk sehari-hari. Peningkatan jumlah penduduk, perkembangan infrastruktur menimbulkan lonjakan permintaan air bersih yang tinggi kebutuhannya dari waktu ke waktu (Sutisna dkk., 2023). Keadaan tersebut memacu penyedia air bersih agar senantiasa mampu memenuhi permintaan masyarakat yang semakin bertambah. Salah satu institusi yang memiliki peran penting dalam hal ini adalah Perusahaan Umum Daerah (PERUMDA) Air Minum (Musyarrof dkk., 2023).

Air menjadi aspek krusial dalam menjaga kelangsungan hidup manusia, karena dipakai untuk berbagai aktivitas harian, termasuk konsumsi, kebersihan, memasak, dan kebutuhan lingkungan. Yang dimaksud dengan air bersih adalah air tawar yang memenuhi standar kualitas, dapat dikonsumsi secara aman, dan tidak berdampak buruk bagi kesehatan manusia (Pahude, 2022). air bersih adalah air tawar yang memenuhi standar kualitas, dapat dikonsumsi secara aman, dan tidak berdampak buruk bagi kesehatan manusia. Namun, berbagai masalah seperti distribusi yang tidak merata, jumlah pengguna yang terus meningkat, serta ketidakakuratan dalam memprediksi penggunaan air sering kali menjadi kendala. Salah satu strategi yang bisa dilakukan adalah melakukan prediksi kebutuhan air pada tahun-tahun berikutnya, agar PDAM dapat memastikan pasokan air yang memadai bagi masyarakat. Selain itu, hasil prediksi kebutuhan air juga dapat dimanfaatkan untuk mengatur alokasi distribusi air kepada pelanggan agar tidak

terjadi kekurangan maupun pemborosan. Prediksi yang akurat akan memberikan manfaat lebih lanjut, yaitu menekan potensi kerugian baik dalam bentuk kehilangan air maupun biaya operasional perusahaan (Fatimah dkk., 2018).

Seiring dengan kemajuan teknologi, algoritma *machine learning* seperti LSTM muncul sebagai salah satu opsi solusi yang efektif. LSTM memiliki kemampuan dalam menangani data serta memberikan kinerja prediksi yang luar biasa. LSTM mampu mengenali pola jangka panjang, menangani data *non-linear*, serta tidak memerlukan asumsi stasioner. Hal ini sangat relevan dengan karakteristik data volume penggunaan air PDAM. Berbagai metode telah dikembangkan dan diuji dalam penelitian terkait prediksi, salah satunya adalah *Long Short-Term Memory* (LSTM). LSTM merupakan salah satu arsitektur dari *Recurrent Neural Network* (RNN) yang dirancang untuk mengolah data berurutan (*sequential data*). Kemampuan ini menjadikan LSTM sangat sesuai digunakan dalam pemodelan dan prediksi data deret waktu (*time series*) (Tita Lattifia dkk., 2022).

Penggunaan algoritma LSTM untuk prediksi konsumsi air pelanggan masih jarang dieksplorasi, meskipun algoritma ini sudah banyak diterapkan di berbagai bidang. Oleh sebab itu, penelitian ini berfokus pada pembangunan model prediksi air berbasis algoritma LSTM. Diharapkan hasil penelitian ini dapat membantu PDAM dalam mengelola penyediaan dan penyaluran air secara lebih efektif serta mendukung pengambilan keputusan berdasarkan informasi data.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana cara membangun model prediksi pemakaian air bersih di Perumda Air Minum menggunakan algoritma LSTM?
2. Seberapa baik kinerja algoritma LSTM dalam memprediksi jumlah penggunaan air bersih berdasarkan metrik evaluasi RMSE dan MAPE?

1.3 Batasan Masalah

Adapun batasan masalah dari penulisan proposal sebagai berikut :

1. Fokus penelitian ini hanya pada prediksi volume pemakaian air bersih dan jumlah pelanggan Perusahaan Umum Daerah Air Minum.
2. Data yang digunakan berasal dari PDAM Kabupaten Kendal, dengan cakupan periode waktu dari tahun 2020 hingga 2025 bulan juni.

1.4 Tujuan

Tujuan penelitian ini adalah menerapkan model *Long Short-Term Memory* (LSTM) untuk memprediksi konsumsi air bersih pada Perusahaan Umum Daerah Air Minum, sehingga dapat mendukung pengelolaan pasokan dan perencanaan distribusi air secara lebih efektif.

1.5 Manfaat

Diharapkan skripsi ini memberikan kontribusi yang bermanfaat secara teori maupun penerapan praktis. Secara teori, penelitian ini mampu memberikan tambahan wawasan serta referensi penerapan *LSTM* untuk prediksi konsumsi air bersih dan berkontribusi pada pengembangan ilmu *machine learning*, prediksi data, dan manajemen air. hasil penelitian ini dapat membantu PDAM dalam mengelola distribusi air bersih secara efisien berdasarkan prediksi penggunaan, sehingga strategi pengelolaan, termasuk kapasitas produksi dan distribusi, dapat dijalankan lebih efektif dan tepat sasaran.

1.6 Sistematika Penulisan

Sistematika penulisan yang akan digunakan oleh penulis dalam sebuah pembuatan laporan tugas akhir adalah sebagai berikut :

- BAB I : Pada bab 1 penulis mengutarakan latar belakang pemilihan judul, rumusan masalah, batasan masalah, tujuan penelitian, metodologi penelitian dan sistematika penulisan.
- BAB II : Pada BAB II memuat tentang penelitian terdahulu dan landasan teori yang berkaitan untuk membantu memahami konsep algoritma LSTM

dalam memprediksi penggunaan air bersih untuk melengkapi penelitian ini.

BAB III : Pada BAB III menjelaskan proses penelitian yang dimulai dari pengumpulan data hingga evaluasi model dari hasil prediksi untuk mendapatkan hasil akurasi prediksi yang baik.

BAB IV : Pada bab IV penulis mengungkapkan hasil prediksi dari penelitian yang menggunakan model *Long Short Term Memory*, serta melakukan evaluasi pengujian streamlit

BAB V : Pada bab V penulis memaparkan kesimpulan dari keseluruhan proses penelitian dari awal hingga akhir serta memberikan saran pengembangan untuk penelitian selanjutnya.



BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Studi sebelumnya yang mengangkat topik *metode Support Vector Regression* dengan *Ant Colony Optimization* untuk memprediksi penggunaan volume air pada PDAM Kota Malang berhasil menghasilkan prediksi yang optimal. Penerapan metode SVR-ACO dalam penelitian itu memberikan akurasi yang baik dengan capaian nilai NMSE 0,281 (Maghfiroh dkk., 2019). Hal ini penggunaan machine learning tepat dapat membantu dalam memperkirakan kebutuhan air secara lebih akurat dan efisien. Berdasarkan hal tersebut, penelitian ini mengajukan penggunaan algoritma LSTM sebagai alternatif yang lebih efisien dan akurat. Dengan memanfaatkan data historis penggunaan air pelanggan, bertujuan untuk membangun sistem prediksi yang handal dan mudah diimplementasikan pada skala operasional PDAM.

Penelitian terdahulu mengenai prediksi produksi padi dengan algoritma *Long Short Term Memory* (LSTM) dilakukan di Kota Lubuklinggau. Studi ini memanfaatkan data produksi padi dari Januari 2019 hingga November 2024 dengan total 71 bulan, dengan pembagian 70% untuk pelatihan dan 30% untuk pengujian. Model terbaik diperoleh dengan konfigurasi *hidden layer* sebanyak 35 neuron, ukuran *batch* 12, dan jumlah *epoch* sebanyak 50. Nilai pengujian MAPE sebesar 4,44% menunjukkan kualitas prediksi yang sangat baik. Model mampu mempelajari pola musiman dengan baik, sehingga hasil prediksi hampir sesuai dengan data aktual (Adhany dkk., 2025).

Penelitian terdahulu yang mengkaji penerapan metode *Extreme Learning Machine* (ELM) dalam memprediksi volume pemakaian air bulanan di Kota Batu menunjukkan bahwa konfigurasi *hyper-parameter* yang paling optimal diperoleh dengan menggunakan 6 *neuron* pada lapisan *input*, 5 *neuron* pada lapisan tersembunyi, serta pembagian data untuk pelatihan dan pengujian sebesar 80%:20%. Penerapan parameter tersebut pada jaringan *Extreme Learning Machine* menghasilkan nilai rata-rata error sebesar 16.437,5 berdasarkan perhitungan *Root*

Mean Square Error (RMSE). Nilai ini mengindikasikan adanya perbedaan rata-rata sebesar 16.437,5 m³ antara data aktual dengan hasil prediksi volume air. Oleh karena itu, metode ELM dinilai kurang optimal dalam kasus ini, yang salah satunya dipengaruhi oleh karakteristik data historis yang digunakan. Penambahan jumlah fitur justru menyebabkan berkurangnya ketersediaan data, sehingga berdampak pada menurunnya kualitas hasil prediksi (Fahrizal dkk., 2021).

Studi sebelumnya mengenai prediksi penggunaan air di PDAM Tirta Dharma Kota Pasuruan berbasis Jaringan Syaraf Tiruan *Backpropagation*. Hasil penelitian menunjukkan bahwa arsitektur 4-2-1, yang melibatkan 4 *neuron* di lapisan *input*, 2 *neuron* di lapisan tersembunyi (*Hidden layer*), dan 1 *neuron* di lapisan output (*output layer*), mampu menghasilkan prediksi dengan performa memuaskan. Untuk prediksi tahun 2023 dengan jumlah iterasi sebanyak 999 epoch, model memperoleh *MSE* 0,001888878 pada data latih dan 0,02720151 pada data uji, serta mencapai akurasi 100%. Untuk tahun 2024, prediksi menggunakan 578 iterasi menunjukkan nilai *MSE* 0,00099987 pada data training dan 0,00053177 pada data testing, dengan akurasi mencapai 100%. Untuk tahun 2025, prediksi dengan 1000 epoch menghasilkan *MSE* 0,00130128 pada data training dan 0,00090659 pada data testing, serta mempertahankan akurasi 100% (Agustina dkk., 2023).

Penelitian terdahulu mengenai prediksi harga saham menggunakan metode RNN dengan pendekatan LSTM menunjukkan hasil pengujian menggunakan indikator MAE dan MAPE, di mana semakin kecil nilai kedua indikator tersebut maka semakin baik performa model dalam menghasilkan prediksi yang akurat. Pada pengujian pertama dengan jumlah epoch sebanyak 10, diperoleh nilai *Mean Absolute Error* (MAE) sebesar 0,0286 serta *Mean Absolute Percentage Error* (MAPE) sebesar 0,0488. Dengan jumlah epoch meningkat menjadi 20, model mencatat nilai MAE 0,0150 dan MAPE 0,0257. Nilai MAE dan MAPE yang rendah menandakan tingkat kesalahan prediksi yang kecil, sehingga model mampu menghasilkan estimasi yang akurat dan mendekati nilai aktual. Dari hasil tersebut, dapat disimpulkan bahwa penggunaan 20 *epoch* pada model LSTM memberikan kinerja prediksi harga saham yang lebih unggul dibandingkan 10 epoch (Hanafiah dkk., 2023).

Hasil penelitian terdahulu menunjukkan bahwa LSTM memberikan kinerja prediksi kualitas udara yang lebih baik dibandingkan metode (*Support Vector Regression*) SVR di Kota Jakarta Pusat. Hal ini dibuktikan melalui hasil peramalan terhadap rata-rata konsentrasi polutan PM10, SO₂, CO, O₃, dan NO₂. Hasil evaluasi dengan MAPE dan RMSE memperlihatkan bahwa model *LSTM* lebih unggul dibandingkan SVR, dengan nilai *MAPE* 12,15% dan *RMSE* 0,0941. Di sisi lain, model SVR mencatat *MAPE* 16,19% dan *RMSE* 0,1666.(Wicaksono dkk., 2023).

Penelitian sebelumnya tentang perbandingan prediksi penggunaan listrik menggunakan metode *Long Short Term Memory* (LSTM) dan *Recurrent Neural Network* (RNN) menunjukkan bahwa hasilnya menunjukkan LSTM lebih baik dalam memprediksi penggunaan listrik dibandingkan RNN. Dalam data siang, kombinasi terbaik pada LSTM diperoleh dengan panjang urutan 20, ukuran hidden 8, tiga lapisan LSTM, serta 70% data pelatihan, dengan rata-rata RMSE sebesar 46,72. Sementara itu, pada data malam, hasil terbaik dicapai dengan panjang urutan 30, ukuran hidden 8, satu lapisan LSTM, dan 80% data pelatihan, dengan rata-rata RMSE mencapai 51,05. Di sisi lain, hasil pengujian RNN menunjukkan bahwa model ini juga bisa membuat prediksi, tapi tingkat keakuratannya lebih rendah. Dengan menggunakan deret data 20 hingga 30, RNN menghasilkan nilai RMSE sekitar 49,18 hingga 49,47 untuk data siang dan 58,68 hingga 57,08 untuk data malam, yang menunjukkan angka lebih tinggi dibandingkan LSTM. Dari sini bisa disimpulkan bahwa meskipun RNN bisa membuat prediksi, LSTM jauh lebih baik karena mampu mengurangi kesalahan dengan lebih signifikan dan lebih efisien dalam menangani deret waktu yang lebih panjang. (Selle, Yudistira and Dewi, 2022).

2.2 Dasar Teori

2.2.1 Air Bersih PDAM

Air bersih adalah kebutuhan dasar bagi manusia dan memiliki peran penting dalam mendukung kehidupan yang berkelanjutan. Oleh sebab itu, penyediaan air bersih perlu ditempatkan sebagai salah satu prioritas utama dalam hal pengelolaan maupun pemenuhan. Ketersediaannya tidak hanya berkaitan erat dengan aspek

kesehatan masyarakat, melainkan juga memberikan dampak besar terhadap kualitas hidup dan tingkat produktivitas sehari-hari. Dalam konteks ini, Perusahaan Daerah Air Minum (PDAM) sebagai lembaga yang bertugas menyediakan layanan air bersih memiliki tanggung jawab penting untuk memastikan kebutuhan masyarakat dapat terpenuhi dengan baik secara berkesinambungan (Tenri dkk., 2022).

Selain itu, air bersih juga memegang peran krusial dalam menunjang berbagai aktivitas harian. Secara umum, pemanfaatan air bersih mencakup kebutuhan dasar seperti penyediaan air minum, keperluan memasak, mencuci, serta berbagai kegiatan domestik lainnya yang mendukung aspek kesehatan dan kelayakan hidup masyarakat (Salman dkk., 2022). Tanpa pasokan air bersih yang lancar, aktivitas rumah tangga maupun kegiatan ekonomi akan sangat terganggu. Perusahaan Daerah Air Minum (PDAM) berfungsi sebagai badan usaha daerah yang bertugas menyediakan layanan air bersih bagi warga.

Tujuan utama PDAM bukan sekadar untuk meraih keuntungan finansial, melainkan lebih ditekankan pada upaya menjamin terpenuhinya hak masyarakat terhadap layanan air minum yang layak, sehat, serta dapat dijangkau oleh seluruh lapisan masyarakat. Sesuai Peraturan Pemerintah Nomor 54 Tahun 2017, PDAM memiliki peran sebagai penyedia layanan publik yang wajib menjamin ketersediaan air minum bagi masyarakat (Gusti dkk., 2019). Secara operasional, PDAM menjalankan dua fungsi. Salah satunya adalah menyediakan air bersih sebagai layanan dasar untuk rumah tangga, usaha, dan industri.

Selain fungsinya sebagai penyedia air bersih, PDAM turut berkontribusi terhadap perkembangan perekonomian daerah. Kontribusi tersebut diwujudkan melalui peningkatan Pendapatan Asli Daerah (PAD), sehingga keberadaan PDAM tidak hanya berfokus pada aspek pelayanan publik, tetapi juga memberikan dampak positif terhadap pembangunan dan kesejahteraan daerah (Fransiska, 2022). Kebutuhan air bersih di setiap daerah tidaklah sama, melainkan bervariasi sesuai dengan karakteristik wilayah masing-masing. Jumlah penduduk, pola penggunaan air, dan tingkat kepadatan wilayah merupakan faktor yang memengaruhi perbedaan tersebut. Seiring berjalannya waktu, permintaan terhadap air bersih cenderung mengalami peningkatan yang signifikan, sedangkan kapasitas pelayanan yang

tersedia tidak selalu dapat menyesuaikan dengan laju pertumbuhan kebutuhan tersebut. Kondisi ini menjadi tantangan bagi PDAM untuk meningkatkan kapasitas dan kualitas layanan agar sesuai dengan kebutuhan di masing-masing daerah.

Peningkatan kebutuhan air bersih dipengaruhi oleh berbagai faktor, di antaranya pertumbuhan jumlah penduduk, meningkatnya tingkat kesejahteraan masyarakat, serta perluasan kawasan perkotaan dan area pelayanan. Di samping itu, perubahan kondisi sosial dan ekonomi masyarakat turut menjadi faktor pendorong, karena semakin tinggi taraf hidup yang dicapai sehingga tingkatnya juga lebih tinggi konsumsi air per kapita yang diperlukan untuk menunjang aktivitas sehari-hari (Carlo dkk., 2024).

PDAM tidak hanya berperan dalam layanan air bersih, tetapi juga berkontribusi pada target *Sustainable Development Goals SDGs* ke-6, yakni penyediaan *Clean Water and Sanitation*. Tujuan ini upaya untuk menjamin ketersediaan air bersih dan layanan sanitasi yang berkelanjutan bagi seluruh lapisan masyarakat. Dalam kerangka tersebut, PDAM diposisikan sebagai faktor utama yang diharapkan mampu memperluas jangkauan akses terhadap air minum yang aman, meningkatkan mutu pelayanan, serta mendorong adanya inovasi teknologi guna menciptakan efisiensi dalam proses distribusi air. Program SDGs mendorong PDAM agar tidak hanya fokus pada penyediaan air, tetapi juga berperan dalam menjaga kelestarian lingkungan, mengurangi kehilangan air, serta mendukung kesehatan masyarakat melalui pelayanan air minum.

2.2.2 Sistem Prediksi

Sistem prediksi adalah suatu mekanisme yang dirancang untuk memperkirakan kondisi, kejadian, ataupun nilai tertentu di masa mendatang dengan memanfaatkan data historis maupun informasi terkini. Prediksi tidak dimaksudkan untuk memberikan hasil yang benar-benar pasti, melainkan berupa estimasi yang sedekat mungkin dengan kenyataan. Dengan demikian, prediksi berperan penting dalam membantu mengantisipasi berbagai bentuk ketidakpastian yang berpotensi terjadi di masa depan. Dengan adanya sistem prediksi, berbagai bidang dapat memperoleh gambaran awal mengenai kecenderungan di masa depan, yang

kemudian digunakan sebagai dasar dalam penyusunan strategi maupun kebijakan (Wijaya Rauf, 2023).

Keberadaan sistem prediksi juga memiliki peran penting dalam mendukung perencanaan jangka pendek maupun jangka panjang. Dalam konteks akademik maupun praktis, sistem ini digunakan untuk menganalisis pola data yang sudah ada, mengenali tren yang sedang berlangsung, dan memproyeksikannya ke periode selanjutnya. Dengan demikian, sistem prediksi bukan hanya berfungsi untuk memberikan angka perkiraan, tetapi juga menjadi sarana analitis yang mampu membantu proses pengambilan keputusan lebih terarah dan terukur (Paradhita, 2024).

Lebih jauh, sistem prediksi umumnya dibangun menggunakan pendekatan statistik maupun algoritma berbasis kecerdasan buatan. Metode statistik seperti regresi linear, ARIMA, dan exponential smoothing sudah lama digunakan untuk memperkirakan data yang memiliki pola tertentu. Di sisi lain, kemajuan teknologi membawa munculnya metode yang lebih modern seperti machine learning dan deep learning, contohnya *Random Forest*, *Support Vector Machine* (SVM), hingga jaringan saraf tiruan (*Artificial Neural Network*) dan *Long Short-Term Memory* (LSTM). Metode ini memungkinkan sistem prediksi untuk menangkap pola yang lebih kompleks, termasuk tren *non-linear*, faktor eksternal, maupun dinamika temporal yang sulit dijangkau oleh metode klasik.

Dalam praktiknya, sistem prediksi telah diterapkan di berbagai bidang seperti ekonomi, kesehatan, transportasi, energi, lingkungan, hingga pendidikan. Misalnya, prediksi kebutuhan energi listrik dapat membantu menjaga kestabilan pasokan, prediksi cuaca mendukung mitigasi bencana, sedangkan prediksi permintaan pasar memungkinkan perusahaan menyusun strategi bisnis yang lebih tepat sasaran. Hal ini menunjukkan bahwa sistem prediksi bukan hanya alat teknis, melainkan juga instrumen strategis yang berkontribusi langsung terhadap efektivitas pengambilan keputusan di berbagai sektor.

2.2.3 *Machine Learning*

Dengan memanfaatkan data prediksi penggunaan air, model dapat belajar prediksi penggunaan air bersih ke masa mendatang. Hal ini memungkinkan sistem untuk melakukan prediksi yang belum pernah dilihat sebelumnya. Penggunaan pendekatan *machine learning* dalam prediksi penggunaan air bersih diharapkan dapat memberikan solusi yang lebih efisien.

Machine learning adalah salah satu bidang dalam kecerdasan buatan yang berfokus pada usaha meniru cara pengambilan keputusan manusia berdasarkan pengalaman yang diperoleh dari data. Prinsip kerjanya adalah menyelesaikan suatu permasalahan atau tugas tanpa harus diberikan instruksi pemrograman yang sangat rinci, melainkan melalui proses pembelajaran dari pola dan pengalaman yang telah ada sebelumnya. Fokus utama adalah mengembangkan sistem yang dapat belajar secara mandiri dalam membuat keputusan, tanpa perlu diprogram ulang, serta mampu beradaptasi dengan perubahan yang ada (Azmi dkk., 2023). Dalam penelitian ini, *machine learning* berperan penting dalam proses pelatihan model prediksi untuk penggunaan air bersih PDAM. Dengan memanfaatkan data historis penggunaan air, model dapat belajar untuk memperkirakan kebutuhan air bersih pada periode mendatang. Hal ini memungkinkan sistem untuk melakukan prediksi terhadap pola yang belum pernah dilihat sebelumnya.

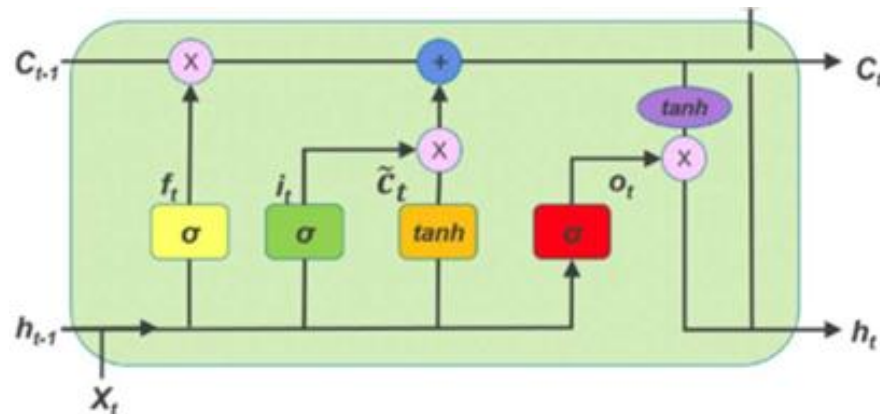
Selain itu, pendekatan *machine learning* juga dapat membantu mengidentifikasi tren, pola musiman, serta potensi anomali yang mungkin terjadi dalam konsumsi air bersih. Dengan demikian, hasil prediksi yang dihasilkan tidak hanya bermanfaat bagi pihak PDAM dalam perencanaan distribusi dan pengelolaan sumber daya, tetapi juga mendukung terciptanya efisiensi operasional serta pelayanan yang lebih baik bagi masyarakat. Penggunaan pendekatan *machine learning* dalam prediksi penggunaan air bersih diharapkan mampu memberikan solusi yang lebih akurat, adaptif, dan berkelanjutan seiring dengan dinamika kebutuhan air yang terus berubah.

2.2.4 Long Short Term Memory

Long Short-Term Memory (LSTM) adalah pengembangan dari teknik *Recurrent Neural Network* (RNN). Algoritma LSTM adalah tipe jaringan saraf buatan yang dibuat khusus untuk mengelola, memproses, dan menganalisis data dalam bentuk deret waktu, contohnya data yang berkaitan dengan volume penggunaan air. Keunggulan model LSTM terletak pada kemampuannya menyimpan informasi dalam periode waktu yang relatif panjang. Hal ini membuat LSTM mampu mengenali serta memahami pola yang lebih kompleks pada data historis.

Dibandingkan dengan metode *time series* konvensional seperti ARIMA atau *Moving Average*, LSTM tidak mengasumsikan data harus stasioner dan lebih unggul dalam menangani data *non-linear* serta memiliki kemampuan belajar pola musiman dan jangka panjang (Panchal dkk., 2024). LSTM juga lebih fleksibel dibanding metode *machine learning* seperti XGBoost atau *Random Forest*, karena dapat memproses urutan waktu secara langsung tanpa perlu rekayasa fitur tambahan yang kompleks. Selain itu, LSTM mampu mengatasi permasalahan *vanishing gradient* melalui penggunaan mekanisme *gates* yang berfungsi untuk mengatur aliran informasi dalam jaringan. Dengan mekanisme tersebut, LSTM dapat mempertahankan serta mengolah informasi dalam jangka panjang tanpa kehilangan data penting pada urutan waktu yang panjang (Febriansyah dkk., 2024).

LSTM memanfaatkan keberadaan sel memori sebagai mekanisme utama untuk mengatasi permasalahan *vanishing gradient*. Struktur jaringan ini terdiri atas beberapa komponen penting, yaitu lapisan *input* (*input layer*), lapisan tersembunyi (*hidden layer*), *cell state*, serta lapisan *output* (*output layer*) yang saling terhubung dalam proses pemrosesan data (Budiprasetyo dkk., 2023).



Gambar 2. 1 arsitektur LSTM

1. Gerbang Masukan (Input Gate)

Pada arsitektur LSTM terdapat sistem gerbang yang berfungsi untuk mengidentifikasi informasi mana yang harus diperbarui. Proses ini dilaksanakan dengan mengonversi nilai ke dalam rentang antara 0 dan 1. Proses itu berlangsung dalam dua tahap. Pertama, lapisan sigmoid berfungsi untuk menentukan informasi mana yang perlu diperbarui (Persamaan 1). Selanjutnya, lapisan \tanh membuat vektor kandidat informasi baru yang akan dimasukkan ke dalam *cell state* (Persamaan 2).

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

Dimana:

i_t : Output dari gerbang masukan pada waktu t

σ : Fungsi aktivasi sigmoid (output antara 0 dan 1)

W_i : Bobot untuk gerbang masukan

h_{t-1}, x_t : Gabungan hidden state sebelumnya dan input sekarang

b_i : Bias untuk gerbang masukan

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (2)$$

Dimana :

\tilde{C}_t : Candidate Cell State

\tanh : Fungsi aktivasi tanh

W_C : Bobot untuk Candidate State

b_C : Bias untuk Candidate State

2. Gerbang *forget* (*forget Gate*)

Pada Pada fase ini, gerbang *forget* berfungsi untuk menentukan informasi mana yang akan dihapus dan informasi mana yang akan tetap disimpan. Proses pengambilan keputusan itu dilaksanakan melalui lapisan sigmoid yang disebut sebagai lapisan gerbang *forget* (Persamaan 3). Lapisan ini menghasilkan keluaran berupa nilai dengan rentang antara 0 hingga 1, yang menentukan seberapa besar informasi dipertahankan atau dihapus.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3)$$

Dimana:

- f_t : Nilai *forget gate* di *timestep* ke- t (antara 0 sampai 1)
- W_f : Bobot *Forget Gate*
- b_f : Bias dari *forget gate*
- σ : Fungsi aktivasi sigmoid (mengubah *output* ke rentang 0–1)
- h_{t-1} : *Hidden state* dari waktu sebelumnya
- x_t : *Input* pada waktu ke- t (saat ini)

3. Status Memori (*Cell State*)

Pada tahap ini, gerbang *forget* memungkinkan penghapusan nilai tertentu pada *cell state* apabila hasil perkalian bernilai mendekati 0 (Persamaan 4).

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

Dimana:

- C_t : *Cell state* (memori jangka panjang) pada waktu ke- t Saat Ini
- C_{t-1} : *Cell State* Sebelumnya (memori dari waktu $t - 1$)
- \tilde{C}_t : Informasi baru (kandidat memori) yang ingin ditambahkan ke *cell state* Gerbang Keluaran (*Output Gate*)
- i_t : *Input gate*, menentukan seberapa besar informasi baru yang akan ditambahkan
- f_t : *Forget gate*, menentukan berapa banyak informasi dari memori lama yang dipertahankan

4. Gerbang Keluaran (*Output Gate*)

Pada tahap ini, gerbang keluaran menentukan keputusan terkait nilai *hidden state* selanjutnya. Perlu diperhatikan bahwa *hidden state* membawa informasi dari masukan sebelumnya. Proses dimulai dengan lapisan sigmoid yang menentukan bagian dari *cell state* yang akan dijadikan *output* (Persamaan 5). Setelah itu, *cell*

state akan diproses menggunakan fungsi aktivasi *tanh* sehingga menghasilkan nilai akhir berupa *hidden state* sebagaimana ditunjukkan pada Persamaan (6). (Kurnia Sari dkk., 2020) .

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

Dimana

o_t : Nilai *output gate* pada waktu ke- t

σ : Fungsi aktivasi sigmoid, menghasilkan nilai antara 0 dan 1

W_o : Bobot (weight) untuk *output gate*

b_o : Bias untuk *output gate*

h_{t-1} : *Hidden state* dari waktu sebelumnya

x_t : *Input* pada waktu ke- t (saat ini)

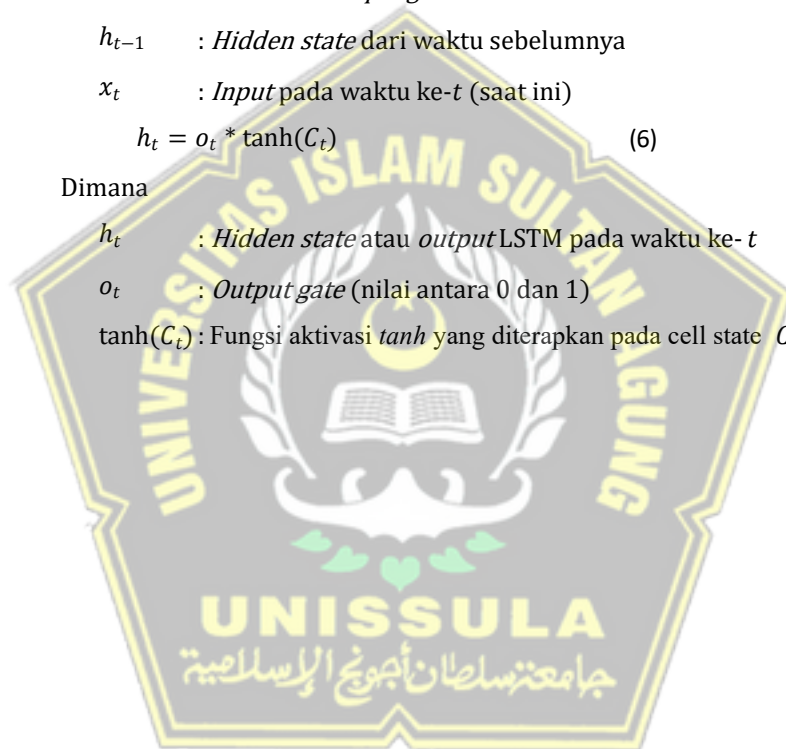
$$h_t = o_t * \tanh(C_t) \quad (6)$$

Dimana

h_t : *Hidden state* atau *output LSTM* pada waktu ke- t

o_t : *Output gate* (nilai antara 0 dan 1)

$\tanh(C_t)$: Fungsi aktivasi *tanh* yang diterapkan pada cell state C_t



BAB III

METODE PENELITIAN

3.1 Metode Penelitian

Penjelasan dalam bagian ini meliputi cara dan tahapan yang diterapkan selama penelitian. Pertama-tama langkahnya yaitu pengumpulan dan pengolahan data, selanjutnya dilakukan *preprocessing* data dan tahap berikutnya pemodelan atau pembuatan model, dan setelah itu melakukan kegiatan pengujian dan hasil.



Gambar 3. 1 *Flowchart* Penelitian

3.1.1 Studi Literatur

Penelitian ini dilakukan dengan mengkaji berbagai sumber referensi, antara lain makalah, jurnal, tesis, serta skripsi terdahulu, disertai dengan penelusuran literatur dari situs web yang relevan. Tujuan dari kajian ini adalah untuk mempelajari teori dan konsep terkait prediksi konsumsi air pelanggan serta memahami penerapan algoritma LSTM.

3.1.2 Pengumpulan Data

Data yang digunakan dalam studi ini bersumber dari dokumentasi penggunaan air oleh pelanggan PDAM. Data tersebut berisi informasi mengenai jumlah konsumsi air pelanggan dalam periode waktu tertentu, misalnya secara bulanan, yang dapat menggambarkan pola penggunaan air. Adapun sumber data diperoleh secara langsung dari instansi PDAM Kabupaten Kendal. Data historis ini digunakan sebagai dasar dalam membangun model prediksi menggunakan algoritma LSTM.

	A	B	C	D	E	F	G	H
1	REKAPITULASI TAGIHAN REKENING AIR							
2	TAHUN 2020							
3								
4			JANUARI			S/D BULAN INI		
5	NAMA	KODE	JUMLAH	JUMLAH	RATA-RATA	JUMLAH	JUMLAH	RATA-RATA
6	CABANG	WIL.	PELANGGAN	PEMAKAIAN	PEMAKAIAN	PELANGGAN	PEMAKAIAN	PEMAKAIAN
7			(SR)	AIR	AIR	(SR)	AIR	AIR
8				(M ³)	(M ³)		(M ³)	(M ³)
9								
10	KENDAL BARAT	02.11	11,580	184,798	15.96	11,580	184,798	15.96
11	KENDAL TIMUR	03.14	8,340	134,690	16.15	8,340	134,690	16.15
12	PATEBON	04.42	4,921	70,280	14.28	4,921	70,280	14.28
13	BRANGSONG	05.13	5,691	86,327	15.17	5,691	86,327	15.17
14	KALI WUNGU I	06.22	8,043	125,469	15.60	8,043	125,469	15.60
15	KALI WUNGU II	07.21	7,931	113,037	14.25	7,931	113,037	14.25
16	WELERI	08.31	6,426	93,031	14.48	6,426	93,031	14.48
17	ROWOSARI	08.32	1,335	17,018	12.75	1,335	17,018	12.75
18	CEPIRING	09.41	9,034	125,643	13.91	9,034	125,643	13.91
19	PEGANDON	10.44	3,741	52,844	14.13	3,741	52,844	14.13
20	SUKOREJO	11.51	4,812	70,858	14.73	4,812	70,858	14.73
21	PATEAN	11.52	1,360	19,088	14.04	1,360	19,088	14.04
22	PAGERUYUNG	12.53	1,235	14,061	11.39	1,235	14,061	11.39
23	BOJA	13.81	5,240	58,557	11.18	5,240	58,557	11.18
24	GEMUH	14.01	3,851	49,314	12.81	3,851	49,314	12.81
25								
26	JUMLAH		83,540	1,215,015	14.54	83,540	1,215,015	14.54
27								
28	JUMLAH	ABT		1,104,640			1,104,640	
29		APT		110,375			110,375	
30								
31								
32								
33								
34								
35								
36								
37								
38								
39								

Gambar 3. 2 Dataset Penggunaan Air

3.1.3 Preprocessing Data

Data yang diperoleh dalam bentuk mentah umumnya belum bisa langsung digunakan untuk membangun model. Karena itu, diperlukan tahapan pra-pemrosesan terlebih dahulu. Langkah ini meliputi beragam prosedur untuk membersihkan dan mempersiapkan data agar siap digunakan dalam pemodelan :

1) Pembersihan Data (*Data Cleaning*):

Proses ini untuk dipastikan data mentah memiliki kualitas yang bisa digunakan. Proses ini mencakup memeriksa dan menangani data yang hilang atau menghapus baris data jika jumlahnya tidak dibutuhkan.

2) Normalisasi Data (*Data Normalization*):

Data tersebut perlu disesuaikan skalanya agar memiliki rentang nilai yang sama, biasanya antara 0 sampai 1. Proses ini disebut normalisasi, dan

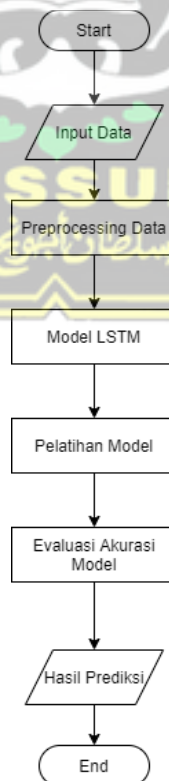
salah satu metode yang umum dipakai adalah *Min-Max Scaling*. Tujuannya supaya model bisa memahami pola data dengan lebih mudah.

3) *Split Data*

Setelah tahap pra-pemrosesan selesai, data kemudian dibagi menjadi dua bagian pada tahap pemisahan data, yaitu data latihan dan data uji dengan perbandingan 80:20. Data latihan digunakan untuk melatih model agar bisa mengenali pola-pola yang terdapat dalam data, sedangkan data uji digunakan untuk mengecek kemampuan model dalam memprediksi data yang belum pernah dilihat sebelumnya. Proses ini sangat penting agar model bisa memberikan hasil prediksi yang akurat dan juga mengurangi kemungkinan terjadinya overfitting.

3.1.4 Model LSTM

Bagian ini menjadi fokus utama dalam penelitian yang bertujuan untuk memprediksi penggunaan air bersih PDAM menggunakan algoritma LSTM. mengidentifikasi pola jangka panjang pada data *time series*, seperti riwayat konsumsi air. Tujuan Tahap ini membangun arsitektur model tersebut.



Gambar 3. 3 Flowchart Perancangan Model

a. *Input Data*

Pada fase ini, dilakukan pengumpulan data historis yang mencakup volume penggunaan air bersih dan jumlah pelanggan aktif setiap bulannya. Kedua data tersebut akan dijadikan sebagai variabel *input* (fitur) dalam pelatihan model prediksi.

b. *Preprocessing Data*

Data yang telah dikumpulkan terlebih dahulu diproses agar siap digunakan oleh model. Langkah awal yang dilakukan adalah membersihkan data dari nilai kosong (*missing value*) maupun data ganda (duplikasi). Selanjutnya, data diubah agar semua nilai memiliki rentang yang sama menggunakan metode *Min-Max Scaling*, sehingga memudahkan model dalam belajar. Kemudian, data diatur dalam bentuk urutan waktu (*sequence*) agar sesuai dengan bentuk masukan yang dibutuhkan oleh model LSTM. Proses ini dilakukan dengan bantuan beberapa *library Python*, seperti *pandas* dan *numpy* untuk mengelola data, serta *MinMaxScaler* dari *sklearn* yang digunakan dalam langkah normalisasi..

Setelah proses preprocessing selesai, data selanjutnya dibagi menjadi dua bagian, yaitu data latih yang mencakup 80% dan data uji yang berisi 20%. Tujuannya adalah agar model tidak hanya dilatih menggunakan sebagian besar data, tetapi juga dapat diuji dengan data yang belum pernah dilihat sebelumnya. Dengan cara ini, proses ini membantu mengurangi kemungkinan terjadinya overfitting dan juga menguji seberapa baik kemampuan model dalam membuat prediksi pada data baru.

c. *Model LSTM*

Model prediksi dibangun dengan menggunakan algoritma *Long Short Term Memory* (LSTM) memanfaatkan library *TensorFlow* dan *Keras*. Arsitektur model dirancang secara bertahap untuk menyesuaikan dengan karakteristik data time series *multivariat*, yaitu volume penggunaan air dan jumlah pelanggan aktif setiap bulan.

Model diawali dengan *layer input* yang menerima data dalam bentuk urutan waktu (*sequence*) dengan struktur tiga dimensi, yaitu (jumlah time

step, jumlah fitur). Sebagai contoh, jika digunakan 3 *time step* ke belakang dan 2 fitur (volume air dan jumlah pelanggan), maka input akan memiliki dimensi (3, 2).

Selanjutnya, data akan diproses oleh *layer* LSTM pertama yang terdiri dari 64 unit memori (*neurons*). *Layer* ini dikonfigurasi dengan *return_sequences=True*, sehingga dapat mengeluarkan seluruh output urutan (*sequence*) dan diteruskan ke *layer* LSTM berikutnya. Setelah itu, ditambahkan *Dropout layer* dengan tingkat *dropout* sebesar 20% (*rate* = 0.2) untuk mencegah *overfitting* dengan cara menonaktifkan secara acak sebagian neuron selama proses pelatihan.

Kemudian, *layer* LSTM kedua dengan 64 unit digunakan untuk melanjutkan pembelajaran terhadap pola data yang lebih kompleks. Pada *layer* ini, *return_sequences = False* karena hanya output akhir dari sekuens yang diperlukan untuk prediksi. Output dari *layer* ini diteruskan ke *layer Dense*, yaitu *layer fully connected* dengan 2 unit *neuron* dan fungsi aktivasi *linear*. *Layer* ini berfungsi menghasilkan output akhir berupa nilai prediksi volume pemakaian air bersih dan jumlah pelanggan. Model dikompilasi menggunakan fungsi *loss Mean Squared Error* (MSE) karena sesuai untuk permasalahan regresi. Untuk proses optimasi digunakan *Adam Optimizer*, yang dikenal efisien dan adaptif dalam mempercepat proses konvergensi.

d. Pelatihan Model

Setelah model selesai dirancang dan dikompilasi, tahap berikutnya adalah melakukan proses pelatihan. Proses pelatihan dilakukan dengan menentukan parameter penting seperti jumlah *epoch*, ukuran *batch* (*batch size*), dan *learning rate*. Selain itu, digunakan juga teknik *EarlyStopping* dengan memantau nilai *validation loss*, sehingga pelatihan dapat dihentikan lebih awal apabila model tidak menunjukkan peningkatan performa setelah beberapa *epoch*. Hal ini bertujuan untuk mencegah *overfitting* serta model dapat mempelajari pola data secara lebih optimal.

e. Evaluasi Akurasi Model

Setelah model selesai dilatih, langkah berikutnya adalah mengevaluasi hasil prediksi yang dikeluarkan. Evaluasi ini bertujuan untuk melihat seberapa jauh perbedaan antara hasil prediksi dengan data yang sebenarnya. Dalam penelitian ini, digunakan dua jenis metrik penilaian, yaitu *Root Mean Square Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE). RMSE adalah cara mengukur rata-rata kesalahan yang dihitung dengan mengkuadratkan kesalahan, lalu dijumlahkan dan dibagi, kemudian diakar untuk menghasilkan angka yang lebih mudah dipahami. Semakin kecil nilai RMSE yang didapat, semakin baik kemampuan model dalam melakukan prediksi. Selain itu, MAPE digunakan untuk mengetahui rata-rata persentase kesalahan prediksi dibandingkan dengan data sebenarnya, di mana nilai MAPE yang rendah menunjukkan hasil prediksi yang lebih akurat. Untuk menghitung RMSE dalam penelitian ini, digunakan fungsi *mean_squared_error* dari library *sklearn.metrics*, lalu diakar dengan bantuan fungsi *numpy.sqrt*. Sementara itu, nilai MAPE dihitung dengan memanfaatkan fungsi *mean_absolute_percentage_error* dari library yang sama.

f. Hasil Prediksi

Model yang telah dilatih kemudian dimanfaatkan untuk melakukan prediksi volume pemakaian air pada bulan berikutnya. Hasil prediksi divisualisasikan dengan grafik Matplotlib untuk memudahkan perbandingan antara data aktual dan hasil prediksinya. Selain ditampilkan dalam bentuk grafik, nilai prediksi juga disajikan dalam bentuk angka di bawah grafik untuk memberikan informasi yang lebih rinci. Visualisasi ini membantu PDAM memahami pola konsumsi air dengan lebih jelas serta dapat dijadikan acuan dalam pengambilan keputusan operasional.

3.1.5 Evaluasi Model

Model dievaluasi untuk mengukur seberapa efektif algoritma LSTM dalam memberikan prediksi yang tepat mengenai pemakaian air bersih PDAM. Pada penelitian ini, metrik evaluasi yang digunakan adalah *Root Mean Square Error*

(RMSE) dan *Mean Absolute Percentage Error* (MAPE) (Cahyani dkk., 2023). *Root Mean Square Error* (RMSE) memberikan penalti lebih besar terhadap kesalahan prediksi yang tinggi, sehingga cocok untuk mendeteksi *outlier* dalam hasil prediksi.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_1 - \hat{y}_1)^2} \quad (7)$$

Dimana :

y_1 : nilai aktual

\hat{y}_1 : nilai hasil prediksi

Σ : menjumlahkan semua eror dari semua data

n : jumlah data yang di evaluasi

Selain RMSE, penelitian ini juga menggunakan *Mean Absolute Percentage Error* (MAPE) sebagai metrik evaluasi. MAPE menghitung rata-rata persentase perbedaan antara nilai yang diprediksi dan nilai yang sebenarnya, sehingga hasil evaluasi lebih mudah dimengerti secara praktis. Nilai MAPE yang kecil menunjukkan bahwa prediksi model hampir sesuai dengan data aktual dengan tingkat kesalahan relatif yang rendah. Penggunaan MAPE sangat bermanfaat karena memberikan perspektif yang intuitif mengenai seberapa besar penyimpangan prediksi terhadap kenyataan dalam bentuk persentase, sehingga dapat dijadikan pertimbangan langsung bagi pengambil keputusan di PDAM dalam menilai keakuratan model.

$$MAPE = \frac{100\%}{n} \left| \frac{y_1 - \hat{y}_1}{y_1} \right| \quad (8)$$

Dimana:

y_1 : nilai aktual

\hat{y}_1 : nilai hasil prediksi

n : jumlah data yang dievaluasi

Selain menampilkan angka hasil evaluasi dengan RMSE dan MAPE, penelitian ini juga menggunakan grafik untuk membandingkan data aktual dengan

hasil prediksi. Grafik ini membantu melihat apakah model sudah mengikuti pola penggunaan air dari waktu ke waktu. Dengan begitu, kinerja model tidak hanya dinilai dari hitungan matematis saja, tetapi juga dari seberapa baik model bisa meniru tren yang ada. Cara ini membuat hasil prediksi lebih mudah dipahami dan bisa langsung dimanfaatkan oleh PDAM dalam mengambil keputusan.

3.1.6 Visualisasi Hasil

Langkah ini bertujuan untuk menampilkan kinerja model *Long Short Term Memory* (LSTM) dalam format grafik dan di bawah grafik agar lebih jelas dipahami. Visualisasi ini tidak hanya menampilkan akurasi model dalam bentuk angka, tetapi juga memberikan gambaran secara intuitif mengenai seberapa baik hasil prediksi penggunaan air bersih dibandingkan dengan data aktual dari Perusahaan Umum Daerah Air Minum.

3.1.7 Rancangan GUI



Gambar 3. 4 Rancangan GUI

Rancangan antarmuka pengguna (GUI) pada gambar tersebut dirancang untuk memudahkan pengguna dalam melihat dan memprediksi volume air di beberapa wilayah. Pada bagian kiri atas terdapat menu dropdown select yang berisi daftar pilihan wilayah seperti Kendal Barat, Kendal Timur, Kaliwungu, dan Patebon, yang memungkinkan pengguna memilih daerah yang ingin dianalisis. Tepat di bawahnya terdapat tombol prediksi waktu 3 bulan yang berfungsi untuk memulai proses perhitungan prediksi volume air selama tiga bulan ke depan berdasarkan wilayah yang dipilih.

Di bagian tengah atas halaman terdapat volume air sebagai penanda utama fungsi aplikasi. Di bawahnya terdapat kotak besar yang diberi label grafik volume air yang akan menampilkan grafik tren atau hasil prediksi volume air, sehingga pengguna dapat memantau perubahan secara visual. Selanjutnya, di bagian bawah grafik terdapat kotak angka hasil yang digunakan untuk menampilkan hasil prediksi dalam bentuk angka, sehingga informasi dapat dipahami secara langsung dan kuantitatif. Tata letak ini dibuat sederhana dan terstruktur, dengan area input berada di sisi kiri dan area output berada di tengah, sehingga alur penggunaan menjadi jelas, yaitu memilih wilayah, menekan tombol prediksi, dan melihat hasil prediksi baik dalam bentuk grafik maupun angka.

3.2 *Software Yang Digunakan*

1. Bahasa Pemrograman

Python

Penelitian ini memanfaatkan bahasa pemrograman *Python* karena dinilai unggul dalam pengolahan data serta penerapan algoritma kecerdasan buatan, khususnya pada *machine learning* dan *deep learning*. *Python* memiliki sintaks yang sederhana dan fleksibel, sehingga memudahkan peneliti dalam melakukan eksplorasi data maupun membangun model prediksi yang kompleks. Keunggulan lain *Python* adalah dukungan komunitas yang luas serta ketersediaan pustaka (*library*) yang beragam untuk kebutuhan analisis data maupun pemodelan.

Dalam penelitian ini, digunakan beberapa pustaka penting yang mendukung analisis data deret waktu. *NumPy* dan *Pandas* berperan dalam pengolahan data, mulai dari manipulasi array multidimensi, transformasi data deret waktu, hingga pembersihan dan pengelolaan data dalam bentuk tabel. Untuk mendukung tahap eksplorasi dan komunikasi hasil analisis, digunakan pustaka *Matplotlib* dan *Seaborn* yang mampu menghasilkan visualisasi grafik dengan tampilan informatif dan mudah dipahami.

Adapun pembangunan dan pelatihan model *Long Short-Term Memory* (LSTM) dilakukan dengan memanfaatkan *TensorFlow* dan *Keras*. *TensorFlow* menyediakan framework komputasi numerik yang efisien,

sedangkan *Keras* memberikan antarmuka yang lebih sederhana dalam perancangan arsitektur jaringan saraf tiruan, pengaturan parameter pelatihan, serta proses evaluasi model. Kombinasi kedua pustaka ini memungkinkan peneliti untuk mengimplementasikan model LSTM secara lebih efektif, baik dari sisi kecepatan komputasi maupun kemudahan penggunaan.

Pemanfaatan *Python* beserta pustaka-pustaka terkait memberikan dasar yang kokoh dalam proses penelitian, mulai dari pengumpulan dan pra-pemrosesan data, pembangunan model prediksi, evaluasi kinerja, hingga penyajian hasil secara grafis. Dengan cara ini, *Python* tidak hanya berfungsi sebagai alat pemrograman, tetapi juga sebagai ekosistem komputasi ilmiah yang mendukung pencapaian tujuan penelitian ini dengan optimal.

2. Perangkat Lunak (*software*)

a. Google Colab

Google Colaboratory, atau dikenal sebagai *Google Colab*, adalah layanan komputasi cloud yang ditawarkan oleh Google. Layanan ini memungkinkan pengguna menjalankan kode Python secara daring tanpa harus menginstal perangkat lunak tambahan di komputer. Colab berbasis *notebook environment* yang mirip dengan *Jupyter Notebook*, sehingga pengguna dapat menulis, mengeksekusi, dan mendokumentasikan kode dalam satu tempat secara interaktif.

Dalam penelitian ini, Google Colab digunakan untuk mengolah data, melatih model LSTM, serta melakukan evaluasi hasil prediksi. Salah satu keunggulan Colab adalah ketersediaan akses ke GPU (*Graphics Processing Unit*) dan TPU (*Tensor Processing Unit*) secara gratis, yang mampu mempercepat proses komputasi, terutama pada pelatihan model deep learning dengan dataset berukuran besar. Hal ini membuat proses eksperimen lebih efisien dibandingkan menggunakan CPU standar pada komputer pribadi.

Selain dukungan komputasi, Google Colab juga memiliki kelebihan dalam integrasi dengan *Google Drive*, sehingga penyimpanan, pemanggilan, dan pembaruan dataset maupun file hasil penelitian dapat dilakukan dengan

mudah dan aman. Fitur kolaborasi yang dimilikinya juga memungkinkan beberapa pengguna bekerja pada satu notebook secara bersamaan, sehingga sangat mendukung riset berbasis tim.

Penggunaan Google Colab dalam penelitian ini tidak hanya memudahkan dari segi teknis, tetapi juga mendukung kesinambungan proses penelitian dengan infrastruktur cloud yang handal, fleksibel, dan dapat diakses kapan saja tanpa perlu perangkat keras khusus.

b. *Visual Studio Code*

Visual Studio Code (VS Code) adalah aplikasi pengedit teks yang ringan namun sangat kuat, sehingga sering digunakan untuk menulis dan mengelola kode program secara lokal di komputer pengguna.. VS Code dipilih karena banyaknya ekstensi pendukung, seperti ekstensi *Python*, *Jupyter*, dan lainnya yang sangat membantu dalam pengembangan proyek berbasis data science.

Visual Studio Code yang digunakan untuk menulis, mengedit, dan menjalankan kode program. Dalam penelitian ini, VS Code dimanfaatkan untuk mengembangkan skrip *Python* yang mencakup tahap pra-pemrosesan data, pelatihan model LSTM, serta evaluasi hasil prediksi. VS Code juga mempermudah pengelolaan struktur proyek penelitian, termasuk penyimpanan dataset, model terlatih, dan file pendukung lainnya dalam satu lingkungan kerja yang terorganisir.

c. *Streamlit*

Streamlit merupakan sebuah framework berbasis Python yang dirancang untuk mempermudah pengembangan antarmuka web interaktif tanpa memerlukan pengetahuan mendalam tentang frontend development. Dengan sintaks yang sederhana, peneliti dapat langsung mengubah hasil analisis data atau model *machine learning* menjadi aplikasi web yang dapat diakses melalui browser.

Dalam penelitian ini, *Streamlit* digunakan untuk menampilkan hasil prediksi penggunaan air bersih serta menyajikan grafik tren pemakaian dan

jumlah pelanggan secara interaktif. Aplikasi ini memungkinkan pengguna memilih wilayah tertentu, menentukan periode prediksi, serta langsung melihat visualisasi data yang dihasilkan model LSTM. Hal ini membuat hasil penelitian tidak hanya berupa angka statis, tetapi juga dapat dieksplorasi secara dinamis oleh pengguna.

Keunggulan lain *Streamlit* adalah tampilannya yang responsif, sederhana, dan mudah digunakan, sehingga hasil penelitian dapat disampaikan dengan cara yang lebih jelas, menarik, dan mudah dipahami, baik oleh pengguna teknis maupun non-teknis. Selain itu, *Streamlit* mendukung integrasi dengan berbagai pustaka *Python* seperti *Matplotlib*, *Seaborn*, dan *Plotly*, yang memperkaya tampilan visualisasi data.

Penggunaan *Streamlit* dalam studi ini berkontribusi dalam menghubungkan hasil komputasi ilmiah dengan penyampaian informasi yang lebih mudah dipahami. Dengan begitu, aplikasi prediksi yang dikembangkan tidak hanya berperan sebagai perangkat analisis, tetapi juga sebagai sarana untuk mengkomunikasikan temuan penelitian yang dapat diakses secara luas dan interaktif.

3. Library dan Framework

a. TensorFlow/Keras

TensorFlow adalah *framework* deep learning yang sangat populer dan powerful, sedangkan *Keras* adalah *library* tingkat tinggi yang berjalan di atas *TensorFlow* dan mempermudah pembuatan serta pelatihan model jaringan saraf tiruan. Dalam penelitian ini, *TensorFlow/Keras* digunakan untuk membangun arsitektur model LSTM yang dapat menangkap pola temporal pada data *time series*. *Keras* menyediakan berbagai lapisan seperti LSTM, *Dense*, dan *Dropout* yang disusun secara berlapis untuk membentuk model yang mampu mempelajari hubungan kompleks antar *variabel*. Selain itu, *Keras* memudahkan pengaturan parameter pelatihan seperti *optimizer*, fungsi loss, serta *callbacks* untuk menghentikan pelatihan secara dini jika model sudah konvergen, sehingga proses pelatihan menjadi lebih efisien dan hasilnya optimal.

b. *Pandas*

Pandas merupakan salah satu *library Python* yang populer untuk pengolahan dan analisis data, terutama data terstruktur dalam bentuk tabel yang dikenal sebagai *DataFrame*. Keunggulan utama *Pandas* adalah kemampuannya dalam mengelola data berukuran besar secara efisien, menyediakan berbagai fungsi untuk memuat, membaca, membersihkan, serta memanipulasi dataset dengan sintaks yang sederhana namun powerful.

Dalam penelitian ini, *Pandas* digunakan untuk membaca data dari file CSV yang menjadi sumber utama dataset. Selanjutnya, dilakukan konversi pada kolom tanggal menjadi tipe *datetime*, yang memudahkan proses analisis data deret waktu (*time series*). *Pandas* juga mendukung proses pembersihan data, seperti menghapus nilai kosong (*missing values*), duplikasi, maupun data yang tidak konsisten, sehingga kualitas data lebih terjamin sebelum masuk ke tahap pemodelan.

Selain itu, *Pandas* memungkinkan data untuk diorganisir berdasarkan wilayah dan periode bulanan, yang relevan dalam konteks penelitian prediksi pemakaian air dan jumlah pelanggan PDAM. Proses agregasi dan pengelompokan (*grouping*) data ini membantu dalam membentuk pola yang lebih jelas sesuai kebutuhan analisis. Data yang sudah terstruktur kemudian diubah menjadi format deret waktu (*time series*), sehingga siap diproses lebih lanjut oleh model LSTM.

c. *NumPy*

NumPy merupakan library utama untuk komputasi numerik di *Python* yang menyediakan struktur data array multidimensi (*ndarray*) dan berbagai fungsi matematika yang cepat serta efisien. Keunggulan utama *NumPy* adalah kemampuannya melakukan operasi vektor dan matriks dalam jumlah besar dengan performa tinggi berkat implementasi berbasis bahasa *C*, sehingga jauh lebih efisien dibandingkan perhitungan menggunakan list standar di *Python*.

Dalam penelitian ini, *NumPy* digunakan untuk melakukan perhitungan numerik serta manipulasi *array* multidimensi yang menjadi format data utama dalam machine learning. *NumPy* berperan dalam mengonversi data mentah ke dalam format numerik yang kompatibel dengan model LSTM. Selain itu, *library* ini juga dimanfaatkan dalam proses transformasi data deret waktu (*time series*), seperti mengubah data historis menjadi bentuk *sequence* (*windowing*), yang memungkinkan model mempelajari pola keterkaitan antarperiode.

NumPy juga berperan dalam proses normalisasi nilai agar data berada pada rentang tertentu, serta penyesuaian bentuk data (*reshape*) sehingga sesuai dengan input yang dibutuhkan oleh LSTM, yakni tiga dimensi: jumlah sampel, jumlah time step, dan jumlah fitur. Tanpa penyesuaian ini, model tidak dapat memproses data dengan benar.

d. *Matplotlib*

Pustaka *Matplotlib* digunakan sebagai alat utama dalam menghasilkan grafik yang menampilkan data dan prediksi secara jelas dan informatif. Penelitian ini memanfaatkan *Matplotlib* untuk menggambarkan hubungan antara data aktual dengan hasil prediksi yang dihasilkan oleh model LSTM. Visualisasi dilakukan pada dua indikator penting, yaitu pemakaian air dan jumlah pelanggan, agar dapat memberikan gambaran menyeluruh tentang performa model dalam menangkap pola deret waktu.

e. *Scikit-learn*

Scikit-learn adalah salah satu library Python yang sering digunakan dalam membuat model machine learning karena menyediakan berbagai alat yang lengkap dan mudah digunakan. Salah satu alatnya adalah *MinMaxScaler*, yang berguna untuk mengubah data agar nilai setiap karakteristik berada dalam rentang tertentu, seperti antara 0 sampai 1. Normalisasi ini penting dalam model LSTM karena jaringan saraf sensitif terhadap skala data; tanpa normalisasi, proses pelatihan bisa terhambat atau menghasilkan prediksi yang kurang akurat.

Selain untuk pra-pemrosesan, *Scikit-learn* juga menyediakan berbagai metode evaluasi model. Dalam penelitian ini, digunakan metrik *Root Mean Squared Error* (RMSE) untuk mengukur besarnya kesalahan rata-rata dalam satuan yang sama dengan data aslinya, serta *Mean Absolute Percentage Error* (MAPE) yang menilai rata-rata persentase kesalahan prediksi. Kombinasi kedua metrik ini memberikan gambaran yang lebih komprehensif mengenai tingkat akurasi model, baik dari sisi deviasi absolut maupun persentase kesalahan.



BAB IV

HASIL DAN ANALISIS PENELITIAN

4.1 Hasil Pengumpulan Data

Tabel 4. 1 Hasil pengumpulan data

bulan	jumlah_pelanggan	jumlah_pemakaian_air	wilayah
1-Jan-20	5240	58557	BOJA
1-Jan-20	5691	86327	BRANGSONG
1-Jan-20	9034	125643	CEPIRING
1-Jan-20	3851	49314	GEMUH
1-Jan-20	8043	125469	KALIWUNGU_I
1-Jan-20	7931	113037	KALIWUNGU_II
1-Jan-20	11580	184798	KENDAL_BARAT
1-Jan-20	8340	134690	KENDAL_TIMUR
1-Jan-20	1235	14061	PAGERUYUNG
1-Jan-20	1360	19088	PATEAN
1-Jan-20	4921	70280	PATEBON
1-Jan-20	3741	52844	PEGANDON
1-Jan-20	1335	17018	ROWOSARI
1-Jan-20	4812	70858	SUKOREJO
1-Jan-20	6426	93031	WELERI

Data penelitian ini bersumber dari catatan historis Perusahaan Umum Daerah Air Minum (PDAM) pada periode Januari 2020 hingga Juni 2025. Dataset tersebut berisi informasi mengenai jumlah pelanggan serta volume pemakaian air bersih di setiap cabang wilayah PDAM. Secara umum, data disusun dalam format yang memuat empat variabel utama, yaitu:

1. Bulan : menunjukkan periode waktu pencatatan dalam format bulan dan tahun.
2. Jumlah Pelanggan : total pelanggan aktif pada periode tertentu di masing-masing wilayah.
3. Jumlah Pemakaian Air : total volume air bersih yang digunakan oleh pelanggan (dalam meter kubik) pada periode tersebut.
4. Wilayah : nama cabang wilayah pelayanan PDAM di Kab Kendal.

Pada awal periode penelitian, yaitu bulan Januari 2020, data mencatat 15 wilayah pelayanan, antara lain Boja, Brangsong, Cepiring, Gemuh, Kaliwungu I,

Kaliwungu II, Kendal Barat, Kendal Timur, Pageruyung, Patean, Patebon, Pegandon, Rowosari, Sukorejo, Dan Weleri. Setiap wilayah memiliki karakteristik jumlah pelanggan dan konsumsi air yang berbeda-beda. Data ini selanjutnya akan digunakan sebagai dasar pemodelan prediksi menggunakan metode *Long Short-Term Memory* (LSTM).

4.2 Hasil *Preprocessing* Data

Tahap ini berfokus pada pembersihan, perubahan, dan penyusunan data mentah agar sesuai dengan kebutuhan analisis atau pemodelan. Langkah ini penting karena membantu memastikan data memiliki kualitas baik, sehingga mendukung proses pelatihan model dengan lebih efektif..

1) Melihat *missing value*

Tahap ini bertujuan untuk memastikan bahwa dataset yang digunakan tidak memiliki nilai yang hilang (*missing value*). Kehadiran data yang hilang dapat menyebabkan bias, menurunkan kualitas analisis, dan memengaruhi kinerja model prediksi. Oleh karena itu, sebelum melanjutkan ke tahap pengolahan dan pelatihan model, dilakukan pemeriksaan *missing data*.

Tabel 4. 2 Data hilang

Nama	Data Hilang
Bulan	0
Jumlah_pelanggan	0
Jumlah_pemakaian_air	0
wilayah	0

Berdasarkan Tabel 4.2, seluruh kolom memiliki nilai *missing value* sebesar **0**, yang berarti tidak terdapat data kosong pada dataset. Dengan kondisi ini, tidak diperlukan proses penanganan *missing data* seperti imputasi atau penghapusan baris.

2) Duplikat data

Selain *missing data*, keberadaan data duplikat juga perlu diperiksa karena dapat menyebabkan bias dan memengaruhi kualitas analisis. Data

duplikat adalah baris data yang memiliki nilai identik pada seluruh kolom dengan baris lainnya. Jika tidak dihapus, duplikasi ini dapat mengakibatkan model prediksi mempelajari pola yang sama berulang kali, sehingga berpotensi menurunkan generalisasi model.

Tabel 4. 3 Data duplikat

Nama	Data duplikat
Bulan	0
Jumlah_pelanggan	0
Jumlah_pemakaian_air	0
wilayah	0

Berdasarkan Tabel 4.3, tidak ditemukan data duplikat pada dataset. Dengan demikian, seluruh baris data yang digunakan adalah unik dan tidak memerlukan proses penghapusan duplikasi.

3) Normalisasi data

Contoh data setelah normalisasi:

	bulan	wilayah	jumlah_pemakaian_air	jumlah_pelanggan
0	2020-01-01	BOJA	0.227991	0.308385
1	2020-01-01	BRANGSONG	0.363295	0.343112
2	2020-01-01	CEPIRING	0.554855	0.600524
3	2020-01-01	GEMUH	0.182956	0.201432
4	2020-01-01	KALIWUNGU_I	0.554007	0.524217

Sebelum data digunakan untuk membangun model, terlebih dahulu dilakukan proses normalisasi. Tujuannya adalah agar skala nilai pada setiap variabel menjadi seragam, sehingga model tidak berat sebelah terhadap variabel yang memiliki nilai lebih besar. Dalam penelitian ini, proses normalisasi diterapkan pada dua variabel utama, yakni jumlah pemakaian air dan jumlah pelanggan. Metode yang digunakan adalah *Min-Max Scaling*, di mana seluruh nilai data diubah ke dalam rentang 0 hingga 1. Proses ini dilakukan menggunakan fungsi *MinMaxScaler* dari *library scikit-learn*. Prinsip kerjanya cukup sederhana, yaitu mengubah setiap nilai menjadi proporsional terhadap nilai terkecil dan terbesar pada variabel tersebut. Dengan cara ini, seluruh nilai data berada pada skala yang sama.

4) Split data

Pada tahap ini, data dibagi menjadi dua bagian, yaitu data latih (*training*) dan data uji (*testing*). Data latih digunakan untuk mengajarkan model agar dapat mengenali pola-pola pada data, sedangkan data uji digunakan untuk melihat seberapa baik model bekerja pada data yang belum pernah dilihat sebelumnya. Dalam penelitian ini, pembagian dilakukan dengan perbandingan 80% untuk data latih dan 20% untuk data uji. Pembagian ini dilakukan dengan menghitung terlebih dahulu jumlah data latih, kemudian memisahkan data fitur dan label berdasarkan indeks tersebut. Data latih diambil dari awal hingga 80% jumlah data, sedangkan sisanya menjadi data uji. Pada penelitian ini, setiap wilayah memiliki 53 data latih dan 13 data uji, dengan total 15 wilayah yang digunakan.

4.3 Implementasi Pelatihan Model

Pada tahap ini, dilakukan pembangunan model *Long Short-Term Memory* (LSTM) untuk memprediksi jumlah pemakaian air dan jumlah pelanggan di setiap wilayah. Model dibangun menggunakan *Sequential API* dari *Keras* dengan struktur jaringan yang terdiri dari dua lapisan LSTM. Lapisan pertama terdiri dari 64 unit neuron dan menggunakan pengaturan *return_sequences=True* sehingga outputnya dapat diteruskan ke lapisan LSTM selanjutnya. Setelah itu ditambahkan lapisan *dropout* dengan tingkat 0,2 untuk mengurangi risiko *overfitting*.

Lapisan kedua merupakan LSTM dengan 64 unit neuron yang diikuti kembali oleh lapisan *dropout* sebesar 0,2. Selanjutnya, digunakan lapisan *Dense* dengan jumlah neuron sesuai banyaknya variabel target, yaitu dua neuron untuk memprediksi jumlah pemakaian air dan jumlah pelanggan secara bersamaan. Model ini menggunakan fungsi aktivasi default pada LSTM serta lapisan keluaran yang bersifat *linear*, sehingga sesuai untuk tugas regresi. Model dikompilasi menggunakan *optimizer Adam* dan fungsi *loss Mean Squared Error* (MSE), yang biasanya digunakan untuk peramalan pada data deret waktu.

Model dilatih dengan menggunakan data yang sudah diatur dalam bentuk urutan, dengan durasi waktu (*time steps*) selama tiga bulan., sehingga setiap tiga bulan data historis digunakan untuk memprediksi nilai pada bulan berikutnya. Data dibagi menjadi dua bagian, yaitu 80% untuk pelatihan (*training set*) dan 20% untuk

pengujian (*testing set*). Parameter pelatihan yang digunakan antara lain jumlah *epoch* maksimum sebanyak 200 dan ukuran *batch* sebesar 16. Untuk mencegah terjadinya overfitting, digunakan teknik *Early Stopping* dengan memantau nilai loss pada data validasi, di mana proses pelatihan akan dihentikan jika tidak terdapat peningkatan selama 15 epoch. Selain itu, pengaturan *restore_best_weights* diaktifkan agar bobot model kembali ke kondisi terbaiknya berdasarkan performa pada data validasi.

Proses pembuatan dan pelatihan model dilakukan secara terpisah untuk setiap wilayah yang terdapat dalam dataset. Pendekatan ini dilakukan dengan menggunakan perulangan (*looping*) pada daftar wilayah unik yang telah diperoleh dari hasil agregasi data. Untuk setiap wilayah, data disaring berdasarkan nama wilayah tersebut, kemudian diurutkan berdasarkan kolom waktu (bulan). Model LSTM yang telah dirancang dibangun ulang dan dilatih menggunakan data wilayah tersebut, sehingga setiap model dapat menangkap pola historis yang spesifik untuk wilayah yang bersangkutan. Pendekatan ini diharapkan dapat meningkatkan akurasi prediksi, karena setiap wilayah memiliki karakteristik konsumsi air dan jumlah pelanggan yang berbeda-beda.

4.4 Evaluasi Model

Evaluasi dilakukan untuk menilai kemampuan model *Long Short-Term Memory* (LSTM) dalam memprediksi jumlah pemakaian air di tiap wilayah. Kinerja model diukur menggunakan dua metrik utama, yaitu *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE).

1. Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) mengukur rata-rata besarnya kesalahan prediksi dalam satuan yang sama dengan data asli, yaitu meter kubik (m^3). Nilai RMSE yang lebih kecil menunjukkan bahwa model mampu menghasilkan prediksi yang lebih dekat dengan nilai aktual.

2. Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) mengukur tingkat kesalahan prediksi dalam bentuk persentase terhadap nilai aktual. Nilai MAPE yang rendah menunjukkan kesalahan relatif yang kecil, sedangkan nilai yang tinggi menandakan kesalahan prediksi yang lebih besar.

Kedua metrik ini digunakan untuk mengevaluasi performa model pada masing-masing wilayah penelitian. Hasil evaluasi disajikan dalam bentuk tabel untuk menunjukkan nilai RMSE dan MAPE pada tiap wilayah. Selain itu, hasil prediksi juga ditampilkan dalam bentuk grafik perbandingan antara nilai riil dan nilai prediksi, sehingga memudahkan pemahaman dan analisis.

1) Pemakaian Air

Tabel 4. 4 Hasil Evaluasi Pemakaian Air

Daerah	RMSE train	RMSE test	MAPE train	MAPE test
Boja	4746.36 m ³	4694.70 m ³	5.02 %	4.22%
Brangsong	5564.61 m ³	5017.47 m ³	5.64%	4.21%
Cepiring	6253.00 m ³	7749.77 m ³	3.96%	4.07%
Gemuh	3466.92 m ³	4699.57 m ³	4.75%	4.51%
Kaliwungu I	5426.33 m ³	4713.20 m ³	3.08%	2.75%
Kaliwungu II	5426.33 m ³	4713.20 m ³	3.08%	2.75%
Kendal Barat	7951.28 m ³	7089.39 m ³	3.21%	2.65%
Kendal Timur	6584.46 m ³	6010.53 m ³	3.75%	3.37%
Pageruyung	1064.89 m ³	1337.12 m ³	6.03%	8.25%
Patean	1927.16 m ³	2181.44 m ³	7.98%	7.72%
Patebon	3364.48 m ³	4423.76 m ³	3.30%	4.29%
Pegandon	3438.68 m ³	5066.04 m ³	4.70%	6.32%
Rowosari	1365.31 m ³	1108.06 m ³	6.26%	4.38%
Sukorejo	4378.64 m ³	5317.16 m ³	4.95%	5.64%
Weleri	4176.72 m ³	4293.69 m ³	3.60%	3.40%

Berdasarkan hasil evaluasi, nilai RMSE pada sebagian besar wilayah menunjukkan konsistensi antara data latih dan data uji. Misalnya, pada wilayah Boja diperoleh nilai RMSE sebesar 4746,36 m³ pada data latih dan 4694,70 m³ pada data uji, sehingga dapat dikatakan model cukup stabil dalam melakukan prediksi. Hal serupa juga terlihat pada wilayah Kaliwungu I dan Kaliwungu II dengan nilai RMSE yang sama pada data latih sebesar 5426,33 m³ dan pada data uji sebesar 4713,20 m³. Meskipun demikian, terdapat beberapa wilayah yang mengalami peningkatan nilai RMSE ketika diuji dengan data baru, seperti pada wilayah Cepiring yang meningkat dari 6253,00 m³ pada data latih menjadi 7749,77 m³ pada data uji. Kondisi ini menunjukkan bahwa model kurang mampu melakukan generalisasi dengan baik pada wilayah tersebut. Sebaliknya, terdapat wilayah yang justru memiliki nilai RMSE lebih rendah pada data uji, seperti Rowosari yang turun dari 1365,31 m³ menjadi 1108,06 m³, menandakan performa model lebih optimal ketika diuji dengan data baru. Secara keseluruhan, nilai RMSE berkisar antara 1000 - 8000 m³, dengan nilai terendah pada wilayah Pageruyung sebesar 1337,12 m³ dan nilai tertinggi pada wilayah Kendal Barat sebesar 7089,39 m³.

Dari sisi MAPE, sebagian besar wilayah menunjukkan tingkat kesalahan relatif yang rendah, yaitu kurang dari 5 persen. Hal ini menandakan bahwa model mampu memberikan hasil prediksi dengan akurasi yang tinggi. Contohnya, pada wilayah Kaliwungu I dan Kaliwungu II nilai MAPE pada data uji hanya sebesar 2,75 persen, sedangkan pada wilayah Kendal Barat sebesar 2,65 persen. Nilai tersebut dapat dikategorikan sangat baik karena kesalahan prediksi relatif kecil dibandingkan dengan konsumsi air aktual. Namun, terdapat beberapa wilayah yang memiliki nilai MAPE lebih tinggi, seperti Pageruyung sebesar 8,25 persen dan Patean sebesar 7,72 persen. Tingginya nilai kesalahan relatif ini kemungkinan disebabkan oleh jumlah pelanggan yang lebih sedikit atau pola pemakaian air yang fluktuatif, sehingga kesalahan kecil dalam prediksi menghasilkan persentase kesalahan yang cukup besar.

Jika dilihat secara menyeluruh, perbedaan nilai RMSE dan MAPE antara data latih dan data uji relatif kecil sehingga dapat disimpulkan bahwa model tidak mengalami *overfitting* secara signifikan. Model mampu menjaga performa ketika diuji dengan data baru, seperti terlihat pada wilayah Weleri yang memiliki nilai MAPE sebesar 3,60 persen pada data latih dan 3,40 persen pada data uji. Hasil ini membuktikan bahwa model memiliki kemampuan generalisasi yang baik. Sebaliknya, wilayah Cepiring dan Pegandon mengalami kenaikan nilai RMSE maupun MAPE pada data uji sehingga performanya sedikit menurun ketika digunakan pada data baru.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model LSTM mampu memberikan prediksi penggunaan air bersih dengan tingkat akurasi yang baik pada sebagian besar wilayah. Mayoritas wilayah memperoleh nilai MAPE di bawah 5 persen, sehingga model dapat diandalkan untuk membantu PDAM dalam memprediksi kebutuhan air. Variasi nilai RMSE antarwilayah juga menunjukkan perbedaan volume konsumsi air, di mana wilayah dengan pemakaian air yang besar cenderung memiliki nilai RMSE lebih tinggi meskipun kesalahan relatif tetap rendah. Dengan demikian, model yang dikembangkan dinilai cukup efektif untuk diterapkan dalam mendukung pengambilan keputusan terkait distribusi air bersih. bahwa seluruh nilai RMSE dan MAPE yang ditampilkan sudah melalui proses denormalisasi, sehingga hasil evaluasi merepresentasikan kondisi sebenarnya dalam satuan volume pemakaian air (m^3).

2) Jumlah Pelanggan

Tabel 4. 5 Hasil Evaluasi Jumlah Pelanggan

Evaluasi	RMSE train	RMSE test	MAPE train	MAPE test
Boja	125.63	55.64	1.42%	0.57%
Brangsong	161.67	37.63	2.14%	0.46%
Cepiring	446.31	102.68	3.57%	0.65%

Gemuh	141.48	28.51	2.29%	0.38%
Kaliwungu I	93.40	30.97	0.75%	0.29%
Kaliwungu II	160.98	56.91	1.47%	0.51%
Kendal Barat	123.79	89.03	0.67%	0.44%
Kendal Timur	50.75	23.44	0.41%	0.20%
Pageruyung	5.72	2.49	0.35%	0.16%
Patean	20.34	5.61	0.83%	0.24%
Patebon	88.45	27.99	1.08%	0.36%
Pegandon	71.50	57.73	1.49%	1.00%
Rowosari	14.51	6.77	0.80%	0.42%
Sukorejo	18.22	12.21	0.29%	0.19%
Weleri	67.28	30.04	0.68%	0.32%

Berdasarkan hasil evaluasi, nilai RMSE pada data uji menunjukkan bahwa kesalahan prediksi jumlah pelanggan relatif kecil di hampir semua wilayah. Misalnya, pada wilayah Boja diperoleh RMSE sebesar 55,64 pelanggan dengan MAPE hanya 0,57 persen. Hal ini menunjukkan bahwa prediksi jumlah pelanggan di wilayah tersebut sangat mendekati kondisi aktual. Demikian juga pada wilayah Brangsong, meskipun nilai RMSE pada data latih mencapai 161,67 pelanggan, hasil pada data uji justru jauh lebih baik dengan RMSE sebesar 37,63 pelanggan dan MAPE hanya 0,46 persen.

Beberapa wilayah lain juga menunjukkan performa yang sangat baik dengan nilai MAPE di bawah 0,5 persen, antara lain Kendal Timur dengan MAPE sebesar 0,20 persen, Pageruyung sebesar 0,16 persen, dan Sukorejo sebesar 0,19 persen. Nilai ini tergolong sangat rendah, sehingga dapat dikatakan prediksi jumlah pelanggan di wilayah tersebut hampir sama dengan kondisi aktual.

Namun, terdapat beberapa wilayah yang memiliki nilai evaluasi sedikit lebih tinggi dibandingkan wilayah lainnya. Contohnya adalah Cepiring dengan nilai RMSE pada data uji sebesar 102,68 pelanggan dan MAPE 0,65 persen, serta Pegandon dengan RMSE sebesar 57,73 pelanggan dan MAPE

1,00 persen. Meskipun demikian, nilai MAPE yang masih berada di sekitar 1 persen tetap menandakan tingkat akurasi prediksi yang baik.

Jika dibandingkan antara data latih dan data uji, sebagian besar wilayah menunjukkan nilai evaluasi yang lebih kecil pada data uji. Kondisi ini menandakan bahwa model tidak mengalami overfitting, bahkan mampu menghasilkan prediksi yang lebih baik ketika digunakan pada data baru. Contohnya pada wilayah Gemuh dengan RMSE turun dari 141,48 pelanggan pada data latih menjadi hanya 28,51 pelanggan pada data uji, serta MAPE yang menurun dari 2,29 persen menjadi 0,38 persen. Hal serupa juga terlihat pada wilayah Patebon dan Weleri, yang menunjukkan perbaikan akurasi pada data uji.

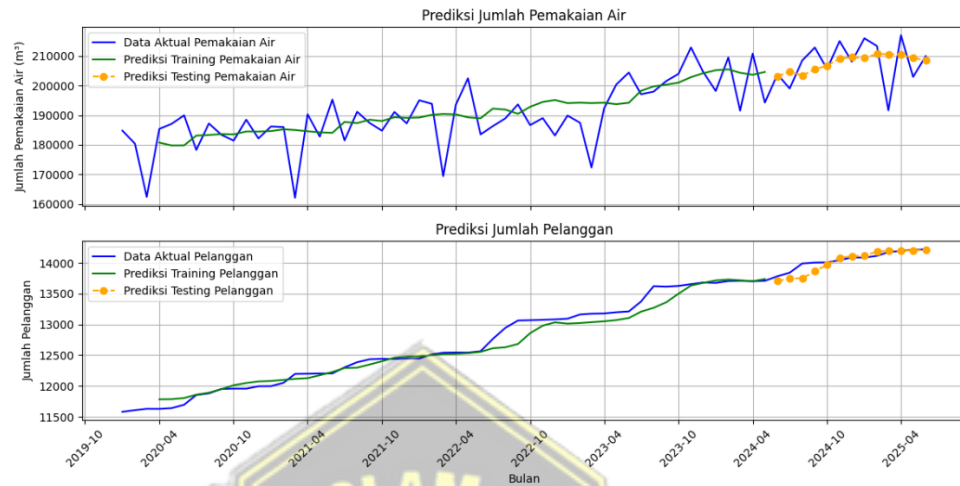
Secara keseluruhan, hasil evaluasi prediksi jumlah pelanggan menunjukkan bahwa model LSTM yang dikembangkan mampu menghasilkan prediksi dengan tingkat akurasi yang sangat tinggi. Hal ini dibuktikan dengan nilai MAPE rata-rata yang berada di bawah 1 persen pada sebagian besar wilayah. Dengan tingkat kesalahan yang sangat rendah ini, model dapat diandalkan untuk mendukung perencanaan dan pengambilan keputusan PDAM, khususnya dalam memperkirakan pertumbuhan jumlah pelanggan di setiap wilayah pelayanan. Perlu dicatat bahwa seluruh nilai RMSE dan MAPE yang ditampilkan sudah melalui proses denormalisasi, sehingga hasil evaluasi merepresentasikan kondisi sebenarnya dalam satuan jumlah pelanggan.

4.5 Visualisasi Hasil

Untuk memudahkan pemahaman mengenai hasil prediksi yang didapatkan, dilakukan visualisasi berupa grafik perbandingan antara data nyata dan data hasil prediksi di beberapa daerah. Tujuan dari visualisasi ini adalah menunjukkan seberapa baik model LSTM mampu mengikuti pola penggunaan air bersih di setiap daerah. Beberapa daerah yang ditampilkan adalah sebagai berikut:

1. Wilayah 1

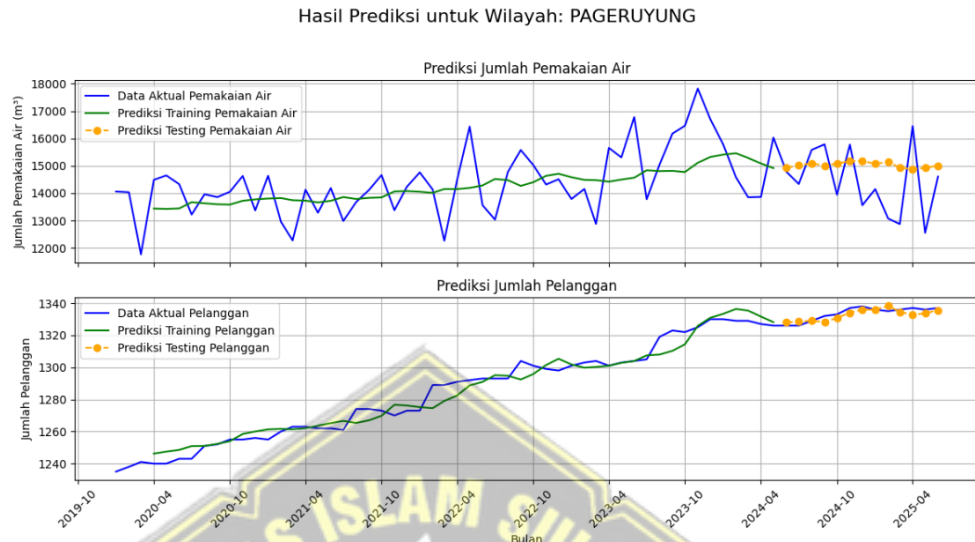
Hasil Prediksi untuk Wilayah: KENDAL_BARAT



Gambar 4. 1 Hasil Visualisasi Wilayah 1

Gambar menunjukkan perbandingan antara data aktual dan hasil prediksi model LSTM di area Kendal Barat untuk dua indikator, yaitu volume pemakaian air dan jumlah pelanggan. Pada indikator penggunaan air, model dapat mengikuti tren dan pola musiman dengan konsisten perbedaan yang muncul umumnya terjadi pada saat lonjakan atau penurunan yang drastis, di mana prediksi cenderung lebih halus dibanding nilai sebenarnya. Sementara itu, untuk indikator jumlah pelanggan, garis prediksi baik dalam data pelatihan maupun pengujian hampir sejajar dengan data aktual, menunjukkan akurasi yang tinggi dan kemampuan generalisasi yang baik. Temuan ini menegaskan bahwa model LSTM ampuh untuk deret waktu yang memiliki tren konsisten, dan tetap mampu menangani deret dengan variasi lebih besar, walaupun penambahan variabel penjelas tambahan bisa meningkatkan akurasi pada titik tertinggi dan terendah nilai.

2. Wilayah 2



Gambar 4. 2 Visualisasi Hasil Wilayah 2

hasil prediksi jumlah pemakaian air menunjukkan bahwa model mampu mengikuti tren rata-rata pemakaian dengan cukup baik, meskipun belum sepenuhnya dapat menangkap fluktuasi yang tinggi pada data aktual. Hal ini terlihat dari nilai RMSE training sebesar 1064,89 m³ dan RMSE testing sebesar 1337,12 m³ dengan tingkat kesalahan relatif MAPE 6,03% pada data training dan 8,25% pada data testing. Nilai tersebut menunjukkan bahwa rata-rata kesalahan prediksi berada pada kisaran 6–8 persen, sehingga model masih cukup layak digunakan untuk memproyeksikan pemakaian air dalam jangka menengah, meskipun pada kondisi ekstrem atau lonjakan konsumsi air model cenderung kurang akurat.

Sementara itu, pada prediksi jumlah pelanggan di wilayah Pageruyung, model menunjukkan performa yang sangat baik. Garis prediksi hampir menempel pada data aktual, baik pada periode training maupun testing. Hal ini didukung oleh hasil evaluasi, di mana diperoleh RMSE training sebesar 5,72 pelanggan dan RMSE testing sebesar 2,49 pelanggan dengan tingkat kesalahan relatif yang sangat rendah, yakni 0,35% pada data training dan 0,16% pada data testing. Nilai kesalahan yang sangat kecil ini mengindikasikan bahwa model memiliki akurasi tinggi dalam

memproyeksikan jumlah pelanggan, sehingga hasil prediksi dapat diandalkan untuk perencanaan kapasitas dan pengembangan layanan di masa mendatang.

4.6 Hasil Implementasi

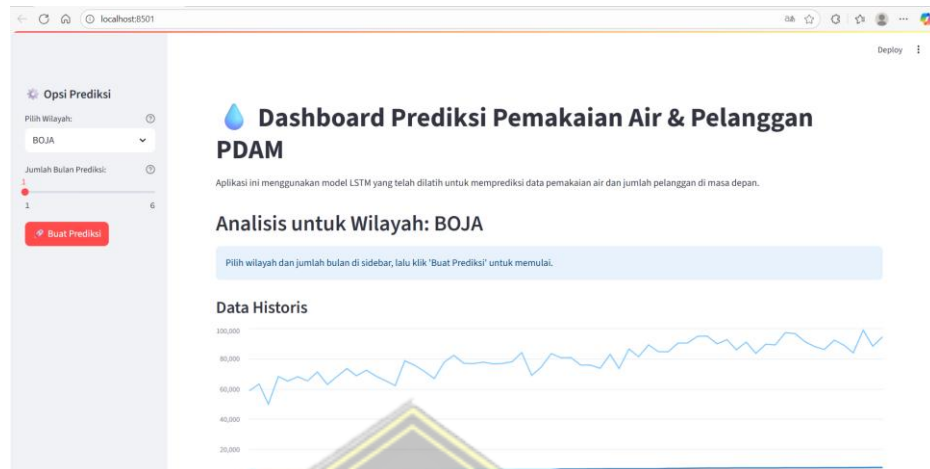
Sebagai tahap akhir implementasi, model prediksi LSTM yang telah dilatih dan dievaluasi di-*deploy* menjadi sebuah aplikasi web interaktif. Langkah ini bertujuan untuk mengubah model dari sekadar kode program menjadi sebuah sistem fungsional yang dapat diakses dan digunakan dengan mudah oleh pengguna. Proses *deployment* dilakukan menggunakan framework Streamlit, sebuah pustaka Python yang memungkinkan pembuatan aplikasi web untuk analisis data dan *machine learning* secara cepat dan efisien.

Tahapannya dimulai dengan menyimpan model LSTM final ke dalam file terpisah agar dapat dimuat kembali saat aplikasi dijalankan. Selanjutnya, dikembangkan skrip utama dengan format `app.py` yang berisi rancangan antarmuka pengguna serta logika sistem. Fitur dalam aplikasi ini mencakup fungsi untuk memuat model LSTM, menerima input data yang relevan (misalnya wilayah dan periode prediksi), memproses data tersebut, menjalankan fungsi prediksi, dan menampilkan hasil berupa angka prediksi serta visualisasi grafik yang membandingkan data historis dengan hasil prediksi.

Aplikasi kemudian dijalankan secara lokal melalui perintah terminal `streamlit run app.py`, yang memungkinkan sistem diakses melalui peramban web di *localhost*. Hasil *deployment* ini adalah sebuah Sistem Prediksi Penggunaan Air Bersih berbasis LSTM yang interaktif dan praktis digunakan, sehingga dapat membantu pihak PDAM dalam memantau tren pemakaian air serta mendukung pengambilan keputusan secara lebih tepat.

Berikut ini merupakan hasil implementasi sistem :

1. Tampilan awal sistem



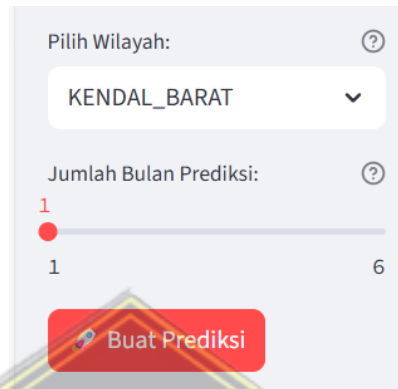
Gambar 4. 3 Tampilan awal sistem

Gambar tersebut menampilkan antarmuka sebuah dasbor aplikasi web yang dirancang untuk Perusahaan Daerah Air Minum (PDAM) dengan tujuan utama melakukan prediksi terhadap pemakaian air dan jumlah pelanggan. Berdasarkan judul dan deskripsi singkatnya, aplikasi ini memanfaatkan teknologi kecerdasan buatan, lebih spesifiknya model *Long Short-Term Memory* (LSTM) dalam menganalisis dan meramalkan data berbasis urutan waktu. Di sisi kiri layar, terdapat panel "Opsi Prediksi" yang berfungsi sebagai pusat kendali bagi pengguna.

Di sini, pengguna dapat secara interaktif memilih wilayah operasional yang spesifik, yang dalam contoh ini adalah "BOJA", serta menentukan horizon waktu prediksi melalui sebuah penggeser, yang saat ini diatur untuk satu bulan ke depan. Setelah parameter diatur, pengguna akan menekan tombol "Buat Prediksi" untuk memulai proses analisis. Area utama dasbor saat ini menampilkan instruksi awal yang memberitahu pengguna untuk melakukan seleksi di panel samping, serta menyajikan visualisasi "Data Historis" untuk wilayah BOJA. Grafik garis pada bagian ini memaparkan data pemakaian air di masa lalu, yang menunjukkan adanya tren peningkatan secara umum meskipun terjadi fluktuasi dari waktu ke waktu. Data historis inilah yang menjadi fondasi bagi model LSTM untuk mempelajari pola dan menghasilkan prediksi yang akurat mengenai permintaan air di masa

mendatang, yang nantinya akan ditampilkan di bagian "Analisis untuk Wilayah" setelah proses prediksi dieksekusi.

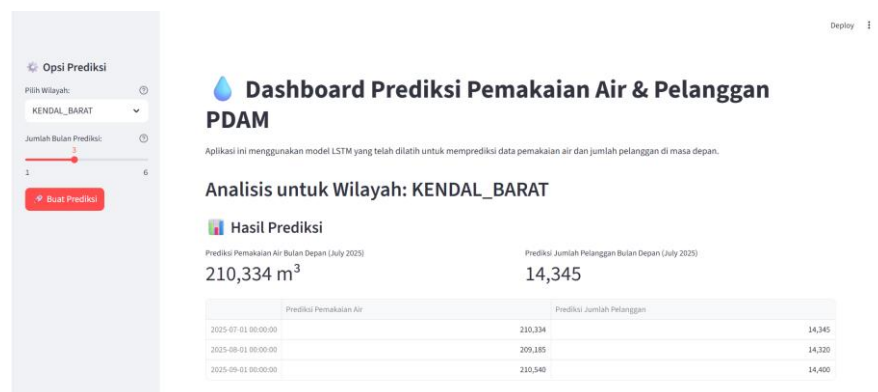
2. Opsi prediksi



Gambar 4. 4 Opsi Prediksi

Gambar tersebut menampilkan sebuah panel kontrol interaktif opsi prediksi yang berfungsi untuk mengatur parameter sebelum menjalankan sebuah peramalan data. Dalam panel ini, pengguna dapat menentukan dua hal utama: pertama, memilih wilayah spesifik yang akan dianalisis melalui menu *dropdown*, di mana "Kendal Barat" saat ini menjadi pilihan. Kedua, pengguna dapat mengatur jangka waktu prediksi yang diinginkan menggunakan penggeser (*slider*) dengan rentang antara satu hingga enam bulan, dan pada gambar ini nilainya diatur untuk satu bulan. Setelah semua parameter diatur, pengguna akan menekan tombol aksi berwarna merah bertuliskan "Buat Prediksi" yang dihiasi ikon roket, yang berfungsi untuk mengeksekusi dan memulai proses prediksi sesuai dengan konfigurasi yang telah dipilih.

3. Hasil Prediksi

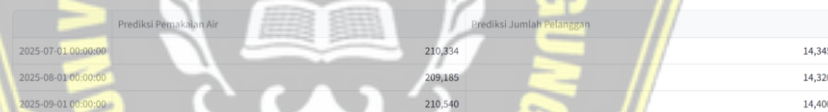


Gambar 4. 5 Hasil Prediksi

Gambar ini menunjukkan tampilan dashboard prediksi pemakaian air dan jumlah pelanggan PDAM untuk wilayah Kendal Barat. Di bagian kiri, wilayah dan jangka waktu prediksi sudah dipilih, yaitu tiga bulan ke depan. Setelah tombol *Buat Prediksi* ditekan, sistem langsung memproses data dan menampilkan hasilnya.

Dari hasil prediksi, pada bulan Juli 2025 pemakaian air di Kendal Barat diperkirakan sekitar 210.334 meter kubik dengan jumlah pelanggan 14.345 orang. Di bulan Agustus, pemakaian air sedikit turun menjadi 209.185 meter kubik dengan jumlah pelanggan 14.320 orang, lalu di bulan September naik lagi menjadi 210.540 meter kubik dengan jumlah pelanggan 14.400 orang. angka-angka ini menunjukkan bahwa pemakaian air di wilayah ini cukup stabil dari bulan ke bulan, sementara jumlah pelanggan perlahan meningkat. Artinya, kebutuhan air tetap tinggi dan ada potensi bertambahnya pengguna layanan PDAM di wilayah tersebut.

4. Tabel

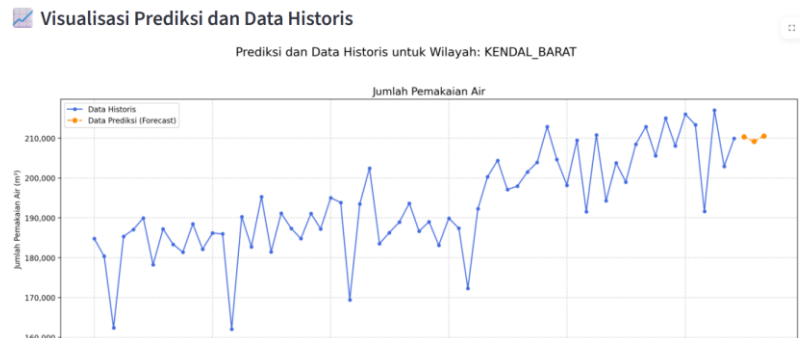


Prediksi Pemakaian Air	Prediksi Jumlah Pelanggan
2025-07-01 00:00:00	210,334
2025-08-01 00:00:00	209,185
2025-09-01 00:00:00	210,540

Gambar 4. 6 Tabel Prediksi

Tabel ini dirancang untuk membantu pengguna memahami prediksi pemakaian air dan jumlah pelanggan PDAM secara lebih praktis. Informasi disajikan dalam format yang rapi sehingga memudahkan pembacaan dan perbandingan antarbulan. Dengan melihat urutan waktu pada kolom pertama, pengguna dapat mengikuti perubahan yang terjadi dari bulan ke bulan, baik dari sisi pemakaian air maupun jumlah pelanggan.

5. Visualisasi Grafik



Gambar 4. 7 Visualisasi grafik

Grafik ini memperlihatkan perbandingan antara data historis pemakaian air dan hasil prediksi untuk wilayah Kendal Barat. Garis berwarna biru menggambarkan catatan pemakaian air pada periode sebelumnya, sedangkan titik berwarna oranye menunjukkan perkiraan pemakaian air untuk beberapa bulan ke depan. Melalui tampilan ini, pengguna dapat dengan mudah melihat pola perubahan penggunaan air dari masa lalu hingga proyeksi di masa depan. Menampilkan data historis dan hasil prediksi dalam satu grafik memberikan visualisasi yang lebih jelas mengenai tren pemakaian air, sehingga memudahkan pengguna untuk memahami kondisi saat ini sekaligus memperkirakan kebutuhan di bulan-bulan mendatang.

6. Pilihan Wilayah

Gambar 4. 8 Pilihan Wilayah

Gambar ini menampilkan menu pilihan wilayah pada aplikasi prediksi pemakaian air dan jumlah pelanggan PDAM. Menu ini berbentuk daftar tarik-

turun (*dropdown*) yang memungkinkan pengguna memilih area yang ingin dianalisis. Fitur ini memudahkan pengguna menyesuaikan analisis berdasarkan lokasi yang relevan dengan kebutuhannya.

Selain itu, menu ini juga berfungsi untuk membatasi ruang lingkup prediksi sehingga hasil yang ditampilkan lebih spesifik dan akurat sesuai dengan karakteristik masing-masing wilayah. Dengan adanya fitur pemilihan wilayah, pengguna dapat membandingkan pola pemakaian antar daerah, memantau tren dari waktu ke waktu, serta menjadikan hasil prediksi sebagai dasar perencanaan distribusi air yang lebih efektif.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Hasil penelitian menunjukkan bahwa algoritma Long Short-Term Memory (LSTM) dapat memberikan prediksi yang akurat mengenai pemakaian air bersih di PDAM Kabupaten Kendal. Hasil evaluasi model menunjukkan nilai RMSE dan MAPE yang cenderung rendah, yang mencerminkan tingkat akurasi prediksi yang tinggi, meskipun di beberapa daerah masih ada hasil yang tidak optimal. Visualisasi perbandingan antara data aktual dan hasil prediksi juga menunjukkan pola yang hampir mirip, yang mengindikasikan bahwa model dapat menangkap tren penggunaan air bersih dengan baik. Ini menunjukkan bahwa penerapan metode deep learning yang berbasis LSTM efektif dan dapat menjadi pilihan solusi dalam sistem prediksi kebutuhan air bersih. Selanjutnya, penelitian ini memberikan kontribusi signifikan dalam mendukung manajemen dan perencanaan distribusi air bersih yang lebih efisien, serta dapat menjadi landasan untuk pengembangan sistem prediksi yang lebih adaptif dan menyeluruh di masa yang akan datang.

5.2 Saran

Berdasarkan temuan penelitian ini, terdapat beberapa saran yang dapat dijadikan acuan untuk pengembangan lebih lanjut.. Pertama, model LSTM yang digunakan masih dapat ditingkatkan akurasinya dengan melakukan *hyperparameter tuning* atau mencoba pendekatan gabungan dengan metode lain, seperti Random Forest atau XGBoost, terutama untuk menghadapi data dengan fluktuasi yang ekstrem. Penelitian ini hanya menggunakan data dari PDAM Kabupaten Kendal, sehingga untuk memperluas cakupan dan menguji kemampuan adaptasi model, disarankan agar penelitian berikutnya mengumpulkan data dari PDAM di wilayah lain yang memiliki kondisi berbeda. Di sisi lain, untuk meningkatkan ketepatan prediksi, diperlukan penambahan variabel pendukung yang lebih beragam, misalnya data cuaca, potensi kebocoran pipa, serta kebijakan

lokal yang dapat memengaruhi penggunaan air. Selain itu, model yang telah dibangun sebaiknya diuji lebih lanjut melalui implementasi langsung pada sistem PDAM yang lebih besar, sehingga dapat dipantau performanya dalam kondisi nyata. Dengan langkah-langkah pengembangan tersebut, diharapkan model prediksi dapat menjadi lebih akurat, andal, dan bermanfaat bagi pengelolaan sumber daya air bersih secara optimal.



DAFTAR PUSTAKA

- Adhany, P. C. *et al.* (2025) 'Prediksi Padi Menggunakan Algoritma Long Short Term Memory', 5(2), pp. 120–127. doi: 10.47065/jimat.v5i2.496.
- Agustina, D. *et al.* (2023) 'Prediksi Distribusi Air Perusahaan Daerah Air Minum (PDAM) Tirta Dharma Kota Pasuruan Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation', *Jurnal PROCESSOR*, 18(1), pp. 8–16. doi: 10.33998/processor.2023.18.1.697.
- Azmi, K., Defit, S. and Sumijan, S. (2023) 'Implementasi Convolutional Neural Network (CNN) Untuk Klasifikasi Batik Tanah Liat Sumatera Barat', *Jurnal Unitek*, 16(1), pp. 28–40. doi: 10.52072/unitek.v16i1.504.
- Budiprasetyo, G., Hani'ah, M. and Aflah, D. Z. (2023) 'Prediksi Harga Saham Syariah Menggunakan Algoritma Long Short-Term Memory (LSTM)', *Jurnal Nasional Teknologi dan Sistem Informasi*, 8(3), pp. 164–172. doi: 10.25077/teknosi.v8i3.2022.164-172.
- Cahyani, J., Mujahidin, S. and Fiqar, T. P. (2023) 'Implementasi Metode Long Short Term Memory (LSTM) untuk Memprediksi Harga Bahan Pokok Nasional', *Jurnal Sistem dan Teknologi Informasi (JustIN)*, 11(2), p. 346. doi: 10.26418/justin.v11i2.57395.
- Carlo, M., Pdam, P. and Makassar, K. (2024) '985 | Page', 2, pp. 985–995.
- Djana, M. (2023) 'Analisis Kualitas Air Dalam Pemenuhan Kebutuhan Air Bersih Di Kecamatan Natar Hajimena Lampung Selatan', *Jurnal Redoks*, 8(1), pp. 81–87. doi: 10.31851/redoks.v8i1.11853.
- Fahrizal, M. A., Adinugroho, S. and ... (2021) 'Prediksi Volume Penggunaan Air Bulanan Kota Batu Menggunakan Metode Extreme Learning Machine (ELM)', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(7), pp. 3078–3086. Available at: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/9478/4282>.
- Fatimah, F., Tejawati, A. and Puspitasari, N. (2018) 'Prediksi Pemakaian Air PDAM Menggunakan Metode Simple Moving Average', *Jurnal Rekayasa Teknologi Informasi (JURTI)*, 2(1), p. 55. doi: 10.30872/jurti.v2i1.1410.
- Febriansyah, F., Sujjada, A. and Sembiring, F. (2024) 'Prediksi Harga Bitcoin Menggunakan Algoritma Long Short Term Memory (LSTM)', *INOVTEK Polbeng - Seri Informatika*, 9(1), pp. 450–459. doi: 10.35314/isi.v9i1.4247.
- Fransiska, F. (2022) 'Assessing The Role of Water-Related Regulations and Actors in The Operation of PDAM (Local Drinking Water Company)', *The Indonesian Journal of Planning and Development*, 7(1), pp. 1–13. doi: 10.14710/ijpd.7.1.1-13.
- Gusti, I. *et al.* (2019) 'Water Availability in the Framework of Environmental Justice: Reconstruction of Municipal Waterworks (Pdam) Regulations', *International Journal of Business, Economics and Law*, 20(4), pp. 51–55.
- Hanafiah, A. *et al.* (2023) 'Penerapan Metode Recurrent Neural Network dengan Pendekatan Long Short-Term Memory (LSTM) Untuk Prediksi Harga Saham', *Bulletin of Computer Science Research*, 4(1), pp. 27–33. doi: 10.47065/bulletincsr.v4i1.321.
- Kurnia Sari, W. *et al.* (2020) 'Klasifikasi Teks Multilabel pada Artikel Berita Menggunakan Long Short-Term Memory dengan Word2Vec', *Masa Berlaku*

- Mulai*, 1(3), pp. 276–285.
- Maghfiroh, A., Widodo, A. W. and Sari, Y. A. (2019) ‘Prakiraan Penggunaan Volume Air PDAM Kota Malang Menggunakan Metode Support Vector Regression dengan Ant Colony Optimization’, *Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(10), pp. 10344–10352. Available at: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/6676/3222>.
- Musyarrof, M. A. and Susanty, A. (2023) ‘Peramalan Volume Produksi Air Bersih Menggunakan Metode Time Series (Studi Kasus: PERUMDAM Purwa Tirta Dharma Kabupaten Grobogan)’, *Industrial Engineering Online Journal*, 12(4). Available at: <https://ejournal3.undip.ac.id/index.php/ieoj/article/view/41339>.
- Pahude, M. S. (2022) ‘Analisis Kebutuhan Air Bersih Di Desa Santigi Kecamatan Tolitoli Utara Kabupaten Tolitoli’, *Jurnal Inovasi Penelitian*, 03(02), pp. 4801–4810. Available at: https://www.google.com/url?sa=i&url=https%3A%2F%2Fstp-mataram.e-journal.id%2FJIP%2Farticle%2Fdownload%2F1744%2F1345&psig=AOvVa wlamkE6LLiImG7Y3YJOPjE&ust=1711081233630000&source=images&cd=vfe&opi=89978449&ved=0CAUQn5wMahcKEwjwp_rzwYSFAxUAAA AAHQAAAAAQBA.
- Panchal, S. A., Ferdouse, L. and Sultana, A. (2024) ‘Comparative Analysis of ARIMA and LSTM Models for Stock Price Prediction’, *27th IEEE/ACIS International Summer Conference on Software Engineering Artificial Intelligence Networking and Parallel/Distributed Computing, SNPD 2024 - Proceedings*, (October), pp. 240–244. doi: 10.1109/SNPD61259.2024.10673919.
- Paradhita, A. N. (2024) ‘Prediksi Inflasi di Indonesia Menggunakan Algoritma Fuzzy dengan Bahasa Pemrograman Python’, *Jurnal Penelitian Inovatif*, 4(2), pp. 457–464. doi: 10.54082/jupin.339.
- Salman, A. B. *et al.* (2022) ‘Pengaruh Distribusi Air Bersih Terhadap Kepuasan Pelanggan di Perusahaan Umum Daerah (Perumda) Air Minum Tirta Danau Tempe Kabupaten Wajo’, *PRECISE : Journal of Economic*, 1(2), pp. 32–51.
- Selle, N., Yudistira, N. and Dewi, C. (2022) ‘Perbandingan Prediksi Penggunaan Listrik dengan Menggunakan Metode Long Short Term Memory (LSTM) dan Recurrent Neural Network (RNN)’, *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(1), pp. 155–162. doi: 10.25126/jtiik.2022915585.
- Sutisna and Yuniar, N. M. (2023) ‘Klasifikasi Kualitas Air Bersih Menggunakan Metode Naïve baiyes’, *Jurnal Sains dan Teknologi*, 5(1), pp. 243–246. Available at: <https://doi.org/10.55338/saintek.v5i1.1383>.
- Tenri, A., Musa, R. and Mallombassi, A. (2022) ‘Analisa Kinerja Sistem Pendistribusian Air PDAM Kota Palopo: Studi Kasus Sumber Air Baku Batupapan’, *Jurnal Konstruksi: Teknik, Infrastruktur dan Sains*, 1(12), pp. 50–61.
- Tita Lattifia, Putu Wira Buana and NI Kadek Dwi Rusjyanthi (2022) ‘Model Prediksi Cuaca Menggunakan Metode LSTM’, *JITTER-Jurnal Ilmiah Teknologi dan Komputer*, 3(1).
- Wicaksono, M. E. S., Arya Sasmita, G. M. and Eka Pratama, I. P. A. (2023) ‘Peramalan Kualitas Udara di Kota Jakarta Pusat dengan Metode Long Short-

Term Memory dan Support-Vector Regression’, *JITTER : Jurnal Ilmiah Teknologi dan Komputer*, 4(1), p. 1576. doi: 10.24843/jtrti.2023.v04.i01.p03.

Wijaya Rauf, B. (2023) ‘Prediksi Penduduk Miskin di Daerah Tertinggal Indonesia dengan Algoritma Prophet’, *Jurnal Ilmu Manajemen Sosial Humaniora (JIMSH)*, 5(2), pp. 116–125. doi: 10.51454/jimsh.v5i2.1024.

