

**PREDIKSI KECENDERUNGAN PENYAKIT BERDASARKAN
KEBUTUHAN OBAT MENGGUNAKAN METODE
*LONG SHORT-TERM MEMORY***

TUGAS AKHIR

Diajukan oleh:

**HILDA SABRINA HASHAR
NIM. 220705008**

Mahasiswa Fakultas Sains dan Teknologi
Program Studi Teknologi Informasi



**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI AR-RANIRY
BANDA ACEH
2026 M/1448 H**

LEMBAR PERSETUJUAN

PREDIKSI KECENDERUNGAN PENYAKIT BERDASARKAN KEBUTUHAN OBAT MENGGUNAKAN METODE *LONG SHORT-TERM MEMORY*

TUGAS AKHIR

Diajukan kepada Fakultas Sains dan Teknologi
Universitas Islam Negeri Ar-Raniry Banda Aceh
Sebagai Salah Satu Beban Studi Memperoleh Gelar Sarjana (S1)
Pada Prodi Teknologi Informasi

Oleh:

HILDA SABRINA HASHAR
NIM. 220705008

Mahasiswa Fakultas Sains dan Teknologi
Program Studi Teknologi Informasi

Disetujui Untuk di Munaqasyahkan Oleh:

Pembimbing I,

Malahayati, M.T

NIP. 198301272015032003

Pembimbing II,

Nurrisqa, M.T

NIP. 199704302025052001

Mengetahui,
Ketua Program Studi Teknologi Informasi

Malahayati, M.T

NIP. 198301272015032003

LEMBAR PENGESAHAN

PREDIKSI KECENDERUNGAN PENYAKIT BERDASARKAN KEBUTUHAN OBAT MENGUNAKAN METODE *LONG SHORT-TERM MEMORY*

TUGAS AKHIR

Telah Diuji Oleh Panitia Ujian Munaqasyah Tugas Akhir
Fakultas Sains dan Teknologi UIN Ar-Raniry Banda Aceh dan Dinyatakan Lulus
Serta Diterima Sebagai Salah Satu Bahan Studi Program Sarjana (S1)
Dalam Program Studi Teknologi Informasi

Pada hari/Tanggal: Rabu, 05 Agustus 2026 M

21 Safar 1448 H

Di Darussalam, Banda Aceh

Panitia Ujian Munaqasyah Tugas Akhir:

Ketua,

Malahayati, M.T

NIP. 198301272015032003

Sekretaris,

Nurrizqa, M.T.

NIP. 199704302025052001

Penguji I,

Dr. Hendri Ahmadian, M.I.M

NIP. 198301042014031002

Penguji II

Rahmad Ade Akbar, S.Pd., M.Pd.

Mengetahui,

Dekan Fakultas Sains dan Teknologi

UIN Ar-Raniry Banda Aceh



Prof. DR. Ir. Muhammad Dirhamsyah, M. T., IPU

NIP. 196210021988111001

LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Hilda Sabrina Hashar
Nim : 220705008
Prodi : Teknologi Informasi
Fakultas : Sains dan Teknologi
Judul : Prediksi Kecenderungan Penyakit Berdasarkan Kebutuhan Obat Menggunakan Metode *Long Short-Term Memory*

Dengan ini menyatakan bahwa dalam penulisan tugas akhir ini, saya:

1. Tidak menggunakan ide orang lain tanpa mampu mengembangkan dan mempertanggungjawabkan;
2. Tidak melakukan plagiasi terhadap naskah karya orang lain;
3. Tidak menggunakan karya orang lain tanpa menyebutkan sumber asli atau tanpa izin pemilik karya;
4. Tidak memanipulasi dan memalsukan data;
5. Mengerjakan sendiri karya ini dan mampu bertanggungjawab atas karya ini.

Bila dikemudian hari ada tuntutan dari pihak lain atas karya saya, dan telah melalui pembuktian yang dapat dipertanggungjawabkan dan ternyata memang ditemukan bukti bahwa saya telah melanggar pernyataan ini, maka saya siap dikenai sanksi berdasarkan aturan yang berlaku di Fakultas Sains dan Teknologi UIN Ar-Raniry Banda Aceh.

Demikian pernyataan ini saya buat dengan sesungguhnya dan tanpa paksaan dari pihak manapun.

Banda Aceh, 05 Agustus 2026

Yang Menyatakan



(Hilda Sabrina Hashar)

ABSTRAK

Nama : Hilda Sabrina Hashar
NIM : 220705008
Program Studi : Teknologi Informasi
Judul : Prediksi Kecenderungan Penyakit Berdasarkan Kebutuhan Obat Menggunakan Metode *Long Short-Term Memory*

Tanggal Sidang : 05 Agustus 2026
Jumlah Halaman : 76 Halaman
Pembimbing I : Malahayati, M.T
Pembimbing II : Nurrizqa, M.T
Kata Kunci : *Long Short-Term Memory (LSTM)*, Prediksi Kecenderungan Penyakit, Penjualan Obat, *Time Series*, Klasifikasi Multikelas

Penjualan obat dapat memberikan gambaran mengenai pola kebutuhan obat yang berkaitan dengan kecenderungan penyakit pada suatu periode. Namun, data transaksi penjualan obat yang bersifat deret waktu perlu diolah dan dianalisis agar pola historisnya dapat dimanfaatkan untuk menghasilkan prediksi. Penelitian ini bertujuan membangun model prediksi kecenderungan penyakit menggunakan metode *Long Short-Term Memory (LSTM)* dengan pendekatan klasifikasi multikelas berdasarkan data transaksi penjualan obat. Data yang digunakan merupakan data transaksi penjualan obat dari Distributor Sakti periode Januari 2023 hingga Desember 2024. Tahapan penelitian meliputi preprocessing data, pengelompokan obat ke dalam enam kategori penyakit, agregasi data secara mingguan, normalisasi, pembentukan *sliding window*, pembangunan model LSTM, serta evaluasi hasil prediksi. Model menggunakan empat *time step* sebagai informasi historis untuk memprediksi kategori penyakit dominan pada periode berikutnya. Arsitektur model terdiri atas satu lapisan LSTM dengan 64 unit dan lapisan *Dense* dengan enam keluaran sesuai dengan jumlah kategori penyakit. Evaluasi model dilakukan menggunakan *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *Confusion Matrix*. Hasil penelitian menunjukkan bahwa model LSTM mampu melakukan prediksi kecenderungan penyakit berdasarkan pola kebutuhan obat dengan tingkat *accuracy* sebesar **83,33%**, dengan 10 dari 12 periode pengujian berhasil diprediksi dengan benar. Hasil tersebut menunjukkan bahwa pola transaksi penjualan obat dapat dimanfaatkan sebagai dasar untuk memprediksi kategori penyakit yang dominan pada periode berikutnya menggunakan pendekatan LSTM.

ABSTRACT

Name : Hilda Sabrina Hashar
Student Number : 220705008
Departement : Information Technology
Title : Prediction of Disease Trends Based on Drug Sales Transaction Data Using the *Long Short-Term Memory* (LSTM) Method

Date : 05 August 2026
Number Of Pages : 76 Pages
Supervisor I : Malahayati, M.T
Supervisor II : Nurrizqa, M.T
Keywords : *Long Short-Term Memory (LSTM)*, Disease Trend Prediction, Drug Sales, *Time Series*, Multiclass Classification.

Drug sales can provide an overview of drug demand patterns associated with disease trends during a particular period. However, drug sales transaction data in the form of time series need to be processed and analyzed so that historical patterns can be utilized to generate predictions. This study aims to develop a disease trend prediction model using the *Long Short-Term Memory* (LSTM) method with a multiclass classification approach based on drug sales transaction data. The data used in this study consist of drug sales transaction data from Distributor Sakti covering the period from January 2023 to December 2024. The research stages include data preprocessing, grouping drugs into six disease categories, weekly data aggregation, normalization, *sliding window* formation, LSTM model development, and prediction evaluation. The model uses four *time steps* as historical information to predict the dominant disease category in the subsequent period. The model architecture consists of one LSTM layer with 64 units and a *Dense* layer with six outputs corresponding to the six disease categories. Model performance was evaluated using *Accuracy*, *Precision*, *Recall*, *F1-Score*, and a *Confusion Matrix*. The results show that the LSTM model is capable of predicting disease trends based on drug demand patterns, achieving an *accuracy* of **83.33%**, with 10 out of 12 testing periods correctly predicted. These results indicate that drug sales transaction patterns can be utilized as a basis for predicting the dominant disease category in the subsequent period using the LSTM approach.

KATA PENGANTAR

Bismillahirrahmanirrahim.

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT, Tuhan semesta alam, yang dengan limpahan rahmat, taufik, dan hidayah-Nya senantiasa menyertai setiap langkah penulis dalam menempuh proses akademik hingga pada tahap penyelesaian tugas akhir ini. Shalawat dan salam semoga senantiasa tercurah kepada junjungan kita Nabi Muhammad Shallallahu ‘Alaihi wa Sallam, beserta keluarga, para sahabat, dan seluruh pengikut beliau yang istiqamah hingga akhir zaman.

Atas izin dan kehendak Allah SWT, penulis dapat menyelesaikan tugas akhir yang berjudul **“Prediksi Kecenderungan Penyakit Berdasarkan Kebutuhan Obat Menggunakan Metode Long Short-Term Memory”**. Tugas akhir ini disusun sebagai salah satu syarat akademik untuk memperoleh gelar Sarjana pada Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Ar-Raniry Banda Aceh.

Perkembangan teknologi informasi, khususnya dalam bidang kecerdasan buatan dan pembelajaran mesin, telah membuka peluang besar dalam pengolahan data berskala besar untuk mendukung pengambilan keputusan di berbagai sektor, termasuk sektor kesehatan. Salah satu permasalahan penting dalam bidang kesehatan adalah keterlambatan dalam mengantisipasi lonjakan kasus penyakit kecenderungan, yang sering kali berdampak pada keterbatasan ketersediaan obat dan kesiapan layanan kesehatan. Oleh karena itu, pemanfaatan data penjualan obat di distributor sebagai indikator awal kemunculan pola penyakit menjadi pendekatan yang relevan dan potensial untuk dikembangkan.

Melalui penelitian ini, penulis berupaya menerapkan metode *Long Short-Term Memory* (LSTM) sebagai salah satu arsitektur deep learning yang unggul dalam pengolahan data deret waktu (time series), guna memprediksi pola kecenderungan penyakit berdasarkan data historis penjualan obat. Diharapkan hasil penelitian ini tidak hanya memberikan kontribusi akademik dalam pengembangan

ilmu pengetahuan di bidang teknologi informasi, tetapi juga dapat menjadi referensi awal bagi pihak terkait dalam mendukung perencanaan dan pencegahan penyakit secara lebih efektif dan berbasis data.

Penulis menyadari sepenuhnya bahwa proses penyusunan tugas akhir ini tidak terlepas dari berbagai tantangan, keterbatasan, serta dinamika yang membutuhkan kesabaran, ketekunan, dan bimbingan dari banyak pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih dan penghargaan yang setinggi-tingginya kepada seluruh pihak yang telah memberikan bantuan, dukungan, serta kontribusi selama proses penyusunan tugas akhir ini, khususnya kepada:

1. Teristimewa dan yang tercinta penulis ucapkan dengan rasa kasih yang tiada tara kepada kedua orangtua penulis tersayang. Kepada cinta pertama dan panutan penulis Ayah Hayardi. Terima kasih atas semua Lelah dan rasa sakit yang selalu ayah rasakan namun ayah tahan agar dapat bekerja keras untuk bisa menjadikan kami anak-anak yang bertumbuh dengan Pendidikan dan kehidupan yang sangat baik, untuk semua usaha ayah agar kami anak-anak ayah tidak merasakan susahnyanya menuntut ilmu sambil bekerja seperti yang pernah ayah rasakan dulu. Kepada Malaikat dan Wanita hebat penulis Ibu Hasridanur. Terimakasih senantiasa memberikan kehangatan dan kekuatan untuk raga kecil ini dan yang selalu memberikan setiap untaian do'a terbaik untuk anak-anaknya, yang tak lagi mengenal sakit ditubuh nya untuk selalu terlihat kuat di depan anak anak nya, untuk setiap sujud mama yang tak pernah henti mengingat dan menyelipkan nama anak-anak nya. Terima kasih yang tak henti untuk setiap keberkahan yang Mama dan Ayah selalu usahakan untuk hidup kami anak-anak Mama dan Ayah.
2. Kepada Abang Tersayang dan terkuat, Hiqal Fajrian Hashar Terima kasih telah menjadi motivasi yang secara diam-diam selalu memberikan pelajaran bagi penulis untuk selalu menjadi kuat dan bertahan. Kepada Kakak Tersayang dan terhebat Hafizah Hashar Terima kasih sudah selalu menjadi penuntun arah jalan penulis, yang selalu mendengar dan memberikan dukungan kepada penulis hingga bisa menyelesaikan setiap masa studi yang penulis tempuh dan terutama bisa menyelesaikan karya tulis ilmiah ini. dan untuk Ayah, Mama, Abang,

Kakak Terima kasih selalu menjadi alasan penulis untuk bertahan sampai saat ini.

3. Kepada Kakak sepupu penulis, Humairah Arigaini, Terima kasih untuk selalu mendengarkan dan memberikan nasehat tanpa pernah menghakimi, untuk setiap warna tawa dan canda yang selalu kita bagi Bersama, untuk tetap berada disamping penulis disaat kesulitan. Terima kasih juga karna sudah merawat, menjaga dan mengjarkan penulis dalam banyak hal.
4. Ibu Malahayati, M.T. selaku Penasehat Akademik dan Pembimbing Proposal yang telah meluangkan waktu, tenaga, pikiran dalam memberikan bimbingan, arahan, serta masukan yang konstruktif sehingga penelitian ini dapat tersusun secara lebih sistematis dan terarah. Serta dengan kesabaran dan perhatian telah memberikan bimbingan serta nasihat kepada penulis sejak awal perkuliahan hingga tahap akhir penyelesaian studi.
5. Ibu Nurrizqa, M.T. selaku Pembimbing Proposal yang telah meluangkan waktu, tenaga, pikiran dalam memberikan bimbingan, arahan, serta masukan yang konstruktif sehingga penelitian ini dapat tersusun secara lebih sistematis dan terarah.
6. Bapak Prof. Dr. Ir. M. Dirhamsyah, M.T., IPU selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Ar-Raniry Banda Aceh.
7. Ibu Malahayati, M.T. dan Bapak Khairan AR, M.Kom. selaku Ketua dan Sekretaris Program Studi Teknologi Informasi yang telah memberikan arahan, kebijakan, serta dukungan akademik selama penulis menempuh pendidikan.
8. Seluruh Bapak dan Ibu dosen Program Studi Teknologi Informasi, yang telah membekali penulis dengan ilmu pengetahuan, wawasan, serta pengalaman akademik yang sangat berharga dan menjadi landasan penting dalam penyusunan penelitian ini.
9. Ibu Cut Ida Rahmadiana, S.Si. selaku staf Program Studi Teknologi Informasi yang telah banyak membantu penulis dalam urusan administrasi akademik dengan penuh kesabaran dan keramahan.
10. Kepada seluruh keluarga besar penulis yang senantiasa memberikan do'a, dukungan, semangat, perhatian, serta motivasi selama proses Pendidikan hingga penyusunan skripsi ini selesai. Dan untuk Kehangatan keluarga yang

selalu penulis dapatkan sehingga bisa kuat untuk menjalani proses perkuliahan dan penulisan karna ilmiah ini.

11. Kepada sahabat tersayang penulis, Resa Raditia yang selalu Bersama penulis sejak tahun 2016 terima kasih selalu membersamai penulis dan selalu jadi tempat penulis berbagi keluh kesah dan Bahagia, menjadi pendengar dan *partner* penulis mengenal banyak hal-hal dan pengalaman baru.
12. Kepada seseorang yang tidak kalah penting atas kehadirannya, yaitu Zakiul Fata sebagai *partner* penulis yang telah menemani dan membersamai hari-hari berat penulis, terima kasih sudah selalu mendengarkan keluh kesah dan ketakutan penulis, terima kasih untuk selalu hadir dan siap siaga Ketika penulis memerlukan bantuan, terima kasih untuk setiap masukan dan saran juga arahan Ketika penulis kesulitan dalam mengerjakan beberapa hal termasuk karya ilmiah ini, terima kasih atas waktu dan banyak hal-hal baik yang diberikan kepada penulis juga sudah menjadi *support system* dan motivator terbaik penulis, terima kasih karna sudah menerima kurang yang selalu penulis takutkan, *tapi diterima siapapun aku kau tangan yang terbuka*.
13. Kepada sahabat seperjuangan tersayang Putri Nabilah Tama, Siti Mariatul Ulfa, Maqhfirah, Munisa, Nurul Mirdawati, Sonia Sabira Gayoara, Zakiul Fata, Arif Maulana, Farras Nabil Rivanza, Sulthan Amirul Abrar, Arief Furqany, yang telah menjadi bagian oenting dalam perjalanan perkuliahan penulis. Terima kasih karna sudah selalu baik dan banyak membantu penulis saat mengalami kesulitan, berbagi cerita, keluh kesah dan Bahagia dalam perjalanan perkuliahan yang tidak mudah ini, juga sudah selalu memberikan pengaruh positif kepada penulis, dan telah berjalan Bersama selama perkuliahan ini semoga persahabatan ini membawa keberkahan bagi semua teman-teman dan tidak terhenti di bangku perkuliahan, tetapi terus terjalin disetiap Langkah selanjutnya, aamiin.
14. Kepada sahabat sahabat penulis, Rulia Nur Izzati, Fitrah Ramanda, Resa Raditia, Fathin Hamami, Mulia Ananda, Tasnim Raidar, Badratun Nafis, Amalina, Raisul Fadhla, Zikril Afzali, M. Tawakkal, Alfatar Syahkur, terima kasih telah menjadi teman-teman penulis dari waktu ke waktu dengan banyak cerita dan pelajaran-pelajaran yang terus dberikan, juga memberikan bantuan,

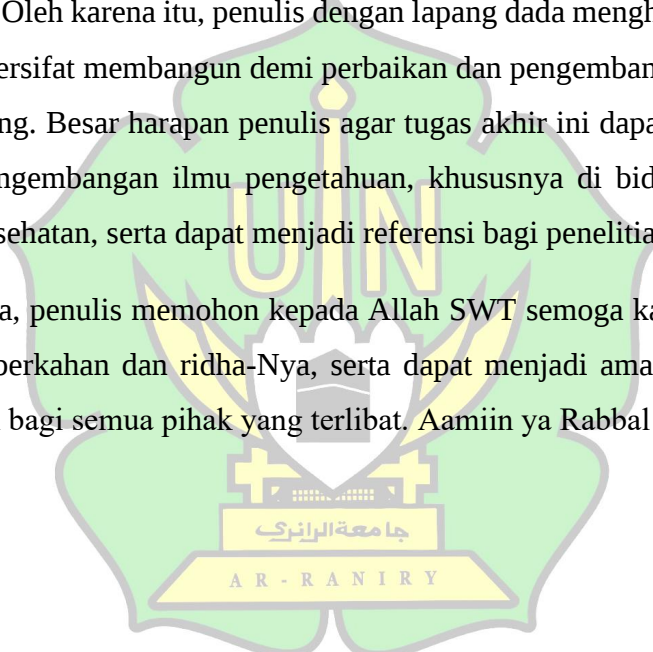
masuk dan saran kepada penulis. Semoga pertemanan kita selalu terjalin terus menerus sampai tua.

15. Kepada seluruh teman-teman Angkatan 2022 Program Studi Teknologi Informasi Universitas Islam Negeri Ar-Raniry Banda Aceh, yang telah menjadi bagian dari perjalanan akademik penulis. Terima kasih atas kebersamaan, kerja sama, serta pengalaman yang telah dibagikan selama masa perkuliahan.
16. Seluruh pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan bantuan, doa, dukungan, motivasi, maupun kontribusi dalam berbagai bentuk selama proses penyusunan skripsi ini. Penulis mengucapkan terima kasih yang sebesar-besarnya.
17. Terakhir, penulis ingin menyampaikan rasa terima kasih yang paling tulus kepada diri sendiri, Hilda Sabrina Hashar. Terima kasih karena sudah berjuang sampai selesai dan tidak pernah benar-benar berhenti, tidak memilih untuk menyerah dan meninggalkan semua yang sudah dimulai. Terima kasih karena sudah berhasil melawan ego, kemalasan dan keinginan untuk berhenti dan menyerah. Terima kasih karena sudah berhasil melawan pikiran-pikiran buruk yang menyiksa diri sendiri. Menjadi anak perempuan yang membawa harapan orang tua dan dengan umur orang tua yang terus menerus menua dan menjadikan itu hal yang sangat dikejar agar bisa segera diselesaikan bukanlah hal yang mudah, ada banyak ketakutan, air mata, rasa lelah yang tidak bisa diceritakan, tidak terlihat dan tidak terbendung, serta kesulitan-kesulitan dan perjuangan yang tidak dilihat orang lain. Namun semua adalah bagian dari proses dan perjalanan yang akan membentuk penulis menjadi pribadi yang kuat, tidak mudah menyerah dan bersyukur untuk setiap hal-hal kecil yang berhasil dilewati. Terima kasih karena masih tetap berusaha meski berkali-kali pernah gagal, tetap bangkit setelah jatuh, tetap berdiri walau terhuyung, tetap mau belajar walau berkali-kali salah, tetap bertahan meski hampir menyerah, terima kasih karena tetap melanjutkan sampai selesai apa yang sudah dimulai. Maaf karena Langkahnya terlalu pelan dan semampunya. Karya ilmiah yang sudah ditulis ini bukan hanya tentang mendapatkan gelar sarjana, namun juga bukti dari setiap usaha, doa, dan kesabaran yang pada akhirnya berhasil di perjuangkan meski dengan Langkah yang pelan dan tertatih-tatih. Semoga semua

yang sudah diperjuangkan senantiasa diiringi kerberkahan Allah SWT. Dan semoga Allah senantiasa memberikan Kesehatan, kemudahan, serta menghadirkan orang-orang baik yang berhati tulus di setiap perjalanan kehidupan. Tidak pernah berhenti berjuang, dan tetap rendah hati serta selalu menjadi ana yang berkati dan pribadi yang bermanfaat bagi keluarga, masyarakat serta orang-orang di sekitar. Terima kasih telah menepati janji pada diri sendiri untuk berjuang dan bertahan sampai selesai, *hancur lebih mudah dari bertahan kupelajari sedari kecil.*

Penulis menyadari bahwa tugas akhir ini masih jauh dari kesempurnaan dan memiliki keterbatasan, baik dari segi cakupan penelitian, metodologi, maupun teknis penulisan. Oleh karena itu, penulis dengan lapang dada mengharapkan kritik dan saran yang bersifat membangun demi perbaikan dan pengembangan penelitian di masa mendatang. Besar harapan penulis agar tugas akhir ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang teknologi informasi dan kesehatan, serta dapat menjadi referensi bagi penelitian selanjutnya.

Akhir kata, penulis memohon kepada Allah SWT semoga karya ilmiah ini mendapatkan keberkahan dan ridha-Nya, serta dapat menjadi amal jariyah yang bernilai kebaikan bagi semua pihak yang terlibat. Aamiin ya Rabbal ‘Alamin.



Banda Aceh, 05 Agustus 2026

Penulis

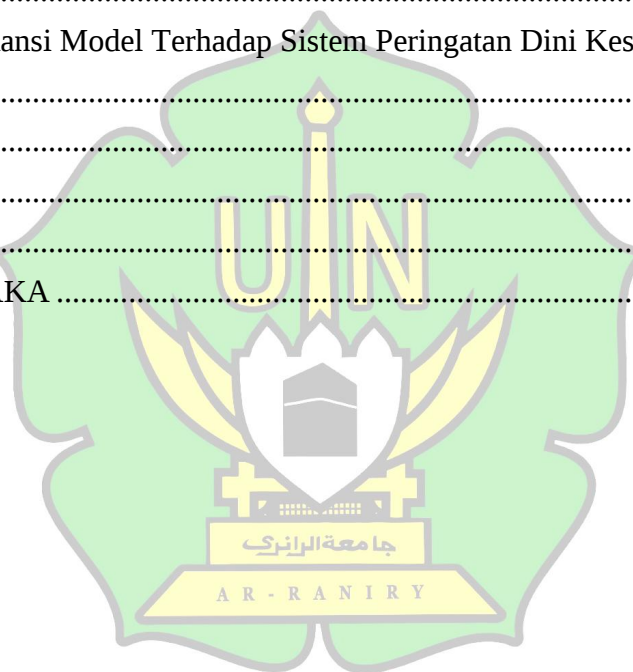
Hilda Sabrina Hashar

DAFTAR ISI

KATA PENGANTAR	i
DAFTAR ISI.....	vii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian.....	4
1.5.1 Manfaat Teoretis	4
1.5.2 Manfaat Praktis	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Penelitian Terdahulu.....	6
2.2 Landasan Teori	8
2.2.1 Prediksi Penyakit	8
2.2.2 Data Penjualan Obat di Distributor sebagai Indikator Penyakit	8
2.2.3 Big Data dalam Kesehatan.....	9
2.2.4 Machine Learning untuk Prediksi Penyakit.....	9
2.2.5 Long Short-Term Memory (LSTM)	9
2.2.6 Sistem Prediksi Penyakit Berbasis LSTM pada Data Distributor	10
2.2.7 Metrik Evaluasi Model Prediksi	10
2.2.8 eXtreme Gradient Boosting (XGBoost)	12
2.2.9 Lingkungan Pengembangan Model (Google Colaboratory).....	13
2.3 Kerangka Berpikir	14
2.4 Hipotesis	16
BAB III METODOLOGI PENELITIAN.....	17
3.1 Jenis Penelitian	17
3.2 Alur Penelitian.....	17
3.3 Populasi dan Sampel	20
3.3.1 Populasi Penelitian.....	20
3.3.2 Sampel Penelitian	21
3.4 Teknik Pengumpulan Data	21

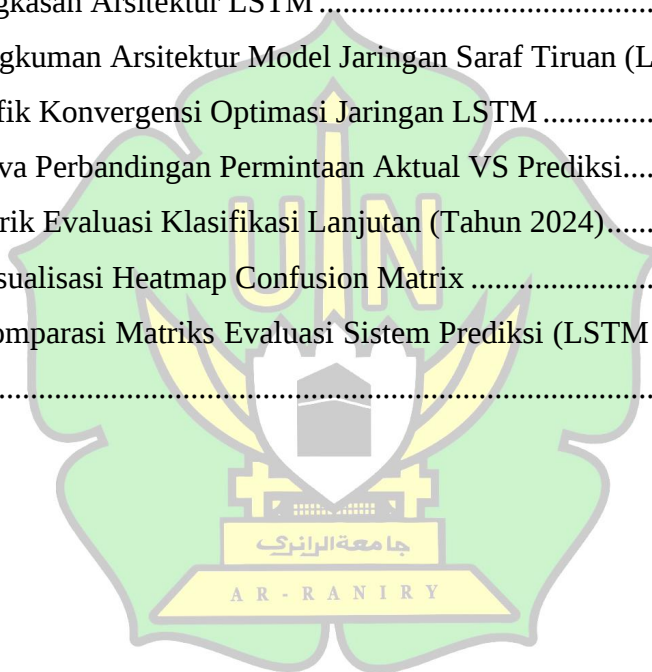
3.4.1 Sumber Data	21
3.4.2 Prosedur Ekstraksi dan Dokumentasi	22
3.5 Teknik Analisis Data	23
3.5.1 Pra-Pemrosesan Data (<i>Data Preprocessing</i>)	23
3.5.2 Transformasi Data Time Series	24
3.5.3 Pembagian Dataset (<i>Training–Testing Split</i>).....	24
3.5.4 Perancangan dan Pelatihan Model LSTM	25
3.5.5 Evaluasi Model Prediksi	25
3.5.6 Interpretasi Hasil dan Visualisasi	26
3.6 Objek dan Subjek Penelitian	26
3.6.1 Objek Penelitian.....	26
3.6.2 Subjek Penelitian	27
3.6.3 Kriteria Subjek dan Objek Data	27
3.7 Lokasi dan Waktu Penelitian.....	27
3.7.1 Lokasi Penelitian.....	27
3.7.2 Waktu Penelitian.....	28
3.8 Perangkat yang Digunakan (Software dan Hardware)	28
3.8.1 Perangkat Lunak (<i>Software</i>)	28
3.8.2 Perangkat Keras (<i>Hardware</i>).....	29
3.9 Variabel Penelitian	30
3.9.1 Variabel Independen (Variabel Bebas).....	30
3.9.2 Variabel Dependen (Variabel Terikat)	30
3.10 Instrumen Penelitian.....	31
BAB IV HASIL DAN PEMBAHASAN	32
4.1 Eksplorasi dan Analisis Data Awal (<i>Exploratory Data Analysis</i>)	32
4.1.1 Profil Dataset dan Pemahaman Atribut	32
4.1.2 Cuplikan Data Mentah	33
4.2 Implementasi Pra-Pemrosesan Data (<i>Data Preprocessing</i>)	34
4.2.1 Ekstraksi dan Kategorisasi Sindrom Penyakit	34
4.2.2 Agregasi Frekuensi Waktu dan Substitusi Nilai Kosong.....	35
4.2.3 Normalisasi Matriks Proporsi Probabilitas	35
4.2.4 Cuplikan Matriks Proporsi Waktu (<i>Snapshot</i>).....	36
4.3 Transformasi Data Deret Waktu Multivariat.....	37
4.3.1 Penggabungan Sinyal Temporal (Time Signal).....	37
4.3.2 Skenario Pembagian Dataset (<i>Train-Test Split</i>).....	38
4.3.3 Restrukturisasi Dimensi <i>Sliding Window</i>	39

4.4 Implementasi dan Pelatihan Model <i>Long Short-Term Memory</i>	41
4.4.1 Konstruksi Arsitektur Jaringan	41
4.4.2 Konfigurasi Fungsi Optimasi dan Kerugian	42
4.4.3 Evaluasi Konvergensi Pelatihan Model	43
4.5 Pengujian dan Hasil Prediksi Sistem	44
4.5.1 Simulasi Prediksi Berjalan (<i>Rolling Prediction</i>)	45
4.5.2 Agregasi Probabilitas Prediksi Bulanan	45
4.5.3 Kalkulasi Tingkat Akurasi Akhir	46
4.5.4 Metrik Evaluasi Lanjutan dan <i>Confusion Matrix</i>	48
4.6 Pembahasan dan Analisis Kritis	51
4.6.1 Interpretasi Hasil Prediksi Terhadap Tren Aktual dan Komparasi Model	51
4.6.2 Signifikansi Model Terhadap Sistem Peringatan Dini Kesehatan	53
BAB V	54
KESIMPULAN	54
5.1 Kesimpulan	54
5.2 Saran	55
DAFTAR PUSTAKA	56



DAFTAR GAMBAR

Gambar 2. 1 Arsitektur LSTM.....	10
Gambar 2. 2 Alur logika rangka pemikiran	15
Gambar 3. 1 Skema Alur Penelitian.....	18
Gambar 4. 1 Proporsi Distribusi Sindrom Penyakit (2023)	36
Gambar 4. 2 Proporsi Distribusi Penyakit (2023).....	37
Gambar 4. 3 Dimensi Tensor Pelatihan	39
Gambar 4. 4 Dimensi Pelatihan Tensor	40
Gambar 4. 5 Ringkasan Arsitektur LSTM	42
Gambar 4. 6 Rangkuman Arsitektur Model Jaringan Saraf Tiruan (LSTM)	42
Gambar 4. 7 Grafik Konvergensi Optimasi Jaringan LSTM	44
Gambar 4. 8 Kurva Perbandingan Permintaan Aktual VS Prediksi.....	48
Gambar 4. 9 Metrik Evaluasi Klasifikasi Lanjutan (Tahun 2024).....	49
Gambar 4. 10 Visualisasi Heatmap Confusion Matrix	50
Gambar 4. 11 Komparasi Matriks Evaluasi Sistem Prediksi (LSTM VS XGBoost)	52



DAFTAR TABEL

Tabel 2. 1 Tabel Penelitian Terdahulu	7
Tabel 3. 1 Deskripsi Atribut Kolom Dataset Mentah Distributor	22
Tabel 3. 2 Spesifikasi Perangkat Lunak	28
Tabel 3. 3 Spesifikasi Perangkat Keras Minimal	29
Tabel 4. 1 Cuplikan Data Mentah Transaksi Distributor (Tahun 2023)	33
Tabel 4. 2 Rincian Distribusi Dataset	38
Tabel 4. 3 Pemilihan Parameter dan Justifikasinya	43
Tabel 4. 4 Laporan Evaluasi Akhir Sistem Prediksi (Tahun 2024)	46



BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan komputasi telah memberikan dampak yang signifikan terhadap sektor kesehatan, terutama dalam hal pengelolaan, penyimpanan, dan pemanfaatan data kesehatan. Pada era digital saat ini, data kesehatan tidak lagi hanya dimanfaatkan sebagai dokumen administratif semata, melainkan telah berkembang menjadi sumber informasi yang bernilai strategis. Data tersebut dapat dianalisis secara sistematis untuk mengidentifikasi pola, kecenderungan, serta dinamika penyebaran penyakit di masyarakat. Melalui pemanfaatan teknik analitik dan pemodelan berbasis data, informasi historis yang tersimpan dalam sistem kesehatan dapat digunakan untuk memperkirakan kemungkinan munculnya penyakit tertentu di masa mendatang (Shahzad et al., 2025).

Salah satu sumber data yang memiliki potensi besar untuk dianalisis dalam konteks kesehatan masyarakat adalah data penjualan obat di distributor. Aktivitas pembelian obat oleh masyarakat umumnya berkaitan erat dengan kondisi kesehatan yang sedang dialami pada periode tertentu, sehingga pola transaksi obat dapat mencerminkan situasi kesehatan masyarakat secara tidak langsung. Ketika terjadi peningkatan kasus suatu penyakit, hal tersebut biasanya diikuti oleh meningkatnya permintaan terhadap obat-obatan yang relevan. Penelitian pada data penjualan di industri farmasi modern menunjukkan bagaimana pendekatan *time series* dan *machine learning* dapat diintegrasikan dalam kerangka kerja yang membantu pengambilan keputusan berbasis data (Ankita et al., 2024).

Secara global, penerapan *big data analytics* dan *machine learning* di sektor kesehatan telah terbukti membantu deteksi dini penyakit, mengidentifikasi pola epidemi, dan memperkirakan kebutuhan farmasi yang memperlihatkan potensi besar dalam mendukung sistem kesehatan (Shah, 2023). Di tingkat nasional, kecenderungan penggunaan obat-obatan mencerminkan dinamika kesehatan

masyarakat. Penelitian oleh Rahmawati dan Lestari (2023) menunjukkan bahwa obat untuk gangguan pencernaan dan pernapasan termasuk yang paling banyak terjual di berbagai distributor di Indonesia. Dengan demikian, data penjualan obat dapat berfungsi sebagai indikator awal (*early indicator*) terhadap munculnya penyakit di suatu wilayah (Wulandari et al., 2022).

Namun, tantangan muncul karena data penjualan obat seringkali belum dimanfaatkan secara optimal. Sebagian besar distributor belum memiliki sistem analitik terintegrasi untuk mengubah data transaksi menjadi informasi prediktif. Pendekatan analisis yang digunakan masih bersifat deskriptif dan belum banyak menerapkan model prediksi berbasis *machine learning* (Supriatna & Permata, 2022). Oleh sebab itu, dibutuhkan metode komputasi yang mampu mengolah data deret waktu (*time-series*) untuk memprediksi kecenderungan penyakit secara lebih akurat.

Salah satu metode yang relevan untuk tugas ini adalah *Long Short-Term Memory* (LSTM), bagian dari algoritma *deep learning* yang dirancang untuk memahami pola data sekuensial. LSTM efektif digunakan untuk memprediksi data yang bergantung pada urutan waktu. Studi perbandingan pada *forecasting* kebutuhan obat menunjukkan bahwa model LSTM umumnya memberikan akurasi lebih baik dan *error* lebih rendah (RMSE/MSE lebih kecil) dibanding ARIMA pada data farmasi (Mbonyinshuti et al., 2024). Beberapa penelitian sebelumnya yang mengelompokkan produk obat menggunakan K-Means dan membangun model Bi-LSTM per klaster juga menunjukkan peningkatan akurasi peramalan permintaan dibanding metode tradisional (Hapsari, 2024; Yusuf et al., 2025).

Meskipun demikian, penelitian yang mengaitkan data farmasi di tingkat distributor lokal dengan kecenderungan penyakit masih jarang ditemukan (Hidayat & Sari, 2022). Sebagian besar studi di Indonesia lebih banyak bersifat operasional persediaan. Maka dari itu, penelitian ini bertujuan untuk mengembangkan model prediksi kecenderungan penyakit menggunakan metode LSTM berbasis data penjualan obat di distributor. Melalui penelitian ini, diharapkan model yang dihasilkan dapat memberikan gambaran yang lebih akurat mengenai tren penyakit

yang mungkin muncul, dan dapat digunakan sebagai alat bantu pendukung keputusan strategis dalam manajemen farmasi maupun kesehatan masyarakat.

1.2 Rumusan Masalah

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

1. Bagaimana cara membangun model prediksi kecenderungan penyakit menggunakan metode Long Short-Term Memory (LSTM) dengan pendekatan klasifikasi multikelas berdasarkan data transaksi penjualan obat?
2. Bagaimana tingkat akurasi model LSTM tersebut dalam memprediksi kategori penyakit yang dominan jika diukur menggunakan metrik evaluasi klasifikasi lanjutan meliputi *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *Confusion Matrix*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, penelitian ini memiliki tujuan untuk:

1. Membangun model prediksi kecenderungan penyakit menggunakan algoritma Long Short-Term Memory (LSTM) untuk mengklasifikasikan kategori penyakit yang dominan berdasarkan pola transaksi penjualan obat.
2. Mengukur dan mengevaluasi tingkat akurasi dari model yang dibangun menggunakan metrik evaluasi klasifikasi yang komprehensif (*Accuracy*, *Precision*, *Recall*, *F1-Score*, *Confusion Matrix*).

1.4 Batasan Masalah

Agar pembahasan lebih terarah dan tidak menyimpang dari tujuan, ruang lingkup penelitian ini dibatasi pada hal-hal berikut:

1. Data primer yang digunakan merupakan rekam data transaksi penjualan obat harian dari distributor farmasi di wilayah Banda Aceh selama periode 1 Januari 2023 hingga 31 Desember 2024.

2. Proses analisis difokuskan secara spesifik pada obat-obatan yang dikelompokkan ke dalam enam kategori sindrom penyakit, yang secara historis memiliki fluktuasi distribusi yang konsisten.
3. Arsitektur model prediksi dibangun menggunakan metode Long Short-Term Memory (LSTM) dengan pendekatan deret waktu multivariat (*multivariate time-series*) untuk memproses fitur masukan berupa proporsi kategori sindrom penyakit dan indeks waktu temporal.
4. Evaluasi performa model prediksi dibatasi murni pada perhitungan matematis tingkat ketepatan prediksi menggunakan metrik evaluasi klasifikasi yang mencakup *Accuracy*, *Precision*, *Recall*, *F1-Score*, serta visualisasi *Confusion Matrix*.
5. Penelitian ini merupakan ranah komputasi berbasis analisis data historis, sehingga tidak menyinggung maupun mengambil kesimpulan mengenai diagnosis medis pasien, penyebab klinis suatu penyakit, dan efektivitas farmakologis dari obat tersebut.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi nyata baik secara teoretis maupun praktis, yang diuraikan sebagai berikut:

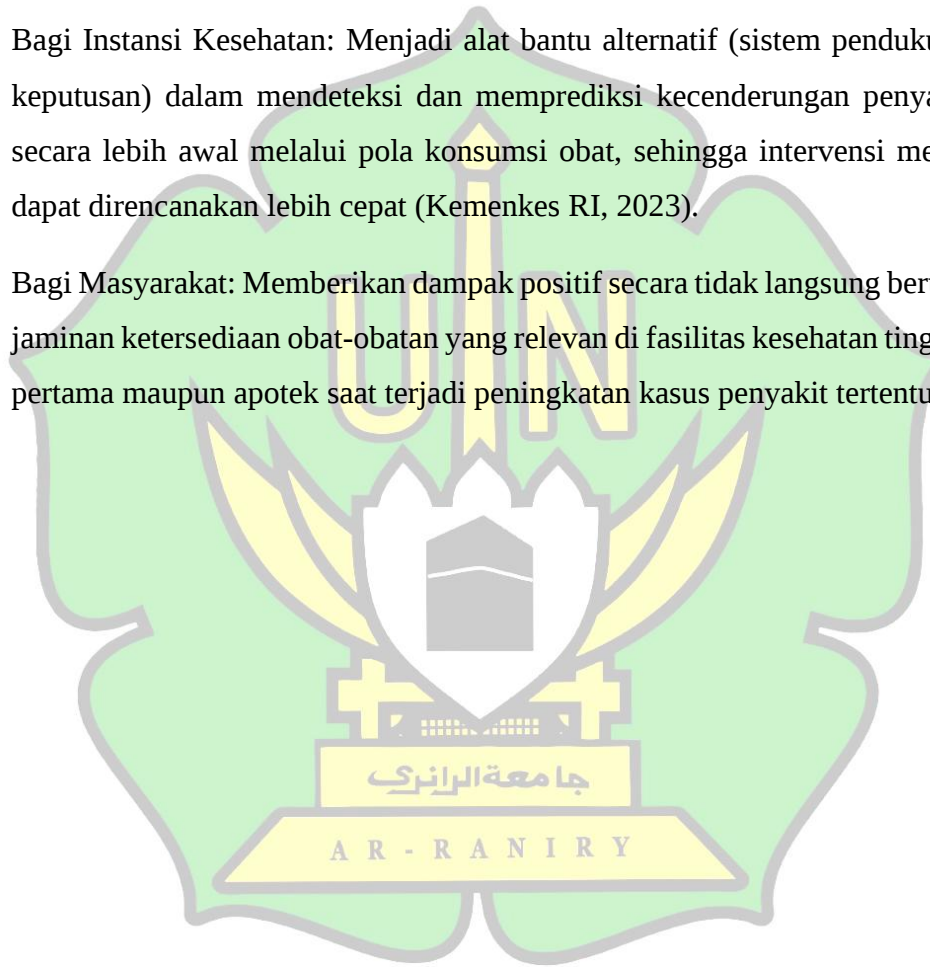
1.5.1 Manfaat Teoretis

Penelitian ini diharapkan dapat memperkaya literatur akademis mengenai penerapan metode *deep learning*, khususnya arsitektur *Long Short-Term Memory* (LSTM) dalam bidang informatika kesehatan. Kajian ini memberikan bukti empiris mengenai pemanfaatan pemrosesan data deret waktu (*time-series*) dari volume distribusi obat sebagai parameter untuk memprediksi kecenderungan penyakit di suatu wilayah. Hasil penelitian ini juga dapat dijadikan referensi dasar bagi studi lanjutan yang ingin mengembangkan sistem surveilans kesehatan berbasis data non-klinis (Setiawan et al., 2021; Putra et al., 2022).

1.5.2 Manfaat Praktis

Secara praktis, hasil dari penelitian dan pemodelan ini memberikan manfaat bagi pihak-pihak terkait, yaitu:

1. Bagi Distributor Farmasi: Membantu manajemen perusahaan dalam mengoptimalkan pengelolaan rantai pasok dan persediaan stok obat secara efisien, sehingga dapat meminimalisasi risiko kelebihan pasokan maupun kelangkaan obat akibat lonjakan permintaan tak terduga.
2. Bagi Instansi Kesehatan: Menjadi alat bantu alternatif (sistem pendukung keputusan) dalam mendeteksi dan memprediksi kecenderungan penyakit secara lebih awal melalui pola konsumsi obat, sehingga intervensi medis dapat direncanakan lebih cepat (Kemenkes RI, 2023).
3. Bagi Masyarakat: Memberikan dampak positif secara tidak langsung berupa jaminan ketersediaan obat-obatan yang relevan di fasilitas kesehatan tingkat pertama maupun apotek saat terjadi peningkatan kasus penyakit tertentu.



BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian sebelumnya telah menyoroti potensi pemanfaatan data penjualan obat untuk mendukung sistem kesehatan masyarakat. Eksplorasi mengenai implementasi *Pharmacy Surveillance Information Systems* (PSIS) menemukan bahwa sistem pada tingkat distributor dapat dijadikan sumber data yang valid untuk surveilans kesehatan publik, meskipun masih terdapat kendala teknis terkait manajemen data dan interoperabilitas (Abbasi et al., 2023). Dalam mendukung sistem tersebut, pemodelan prediktif menggunakan algoritma *machine learning* terus dikembangkan untuk mendapatkan akurasi yang lebih baik. Studi perbandingan pemodelan rantai pasok kesehatan di Afrika menunjukkan bahwa metode *Long Short-Term Memory* (LSTM) memberikan tingkat akurasi yang lebih baik dalam menangkap fluktuasi dan lonjakan permintaan dibandingkan dengan metode statistik konvensional seperti ARIMA (Mbonyinshuti et al., 2024).

Di Indonesia, penerapan arsitektur LSTM untuk memprediksi penjualan obat di jaringan distributor juga telah dilakukan. Pengujian terhadap lima jenis obat menunjukkan bahwa LSTM menghasilkan nilai *Root Mean Square Error* (RMSE) yang lebih rendah dibandingkan ARIMA, kendati model yang dibangun masih menunjukkan indikasi *overfitting* pada beberapa skenario (Kurniawati et al., 2023). Penelitian lain pada skala industri farmasi lokal mengombinasikan metode K-Means untuk pengelompokan data penjualan dengan *Bidirectional* LSTM (Bi-LSTM). Hasilnya menunjukkan bahwa pendekatan klasterisasi tersebut mampu meramalkan permintaan obat dengan tingkat *Mean Absolute Error* (MAE) dan *Mean Squared Error* (MSE) yang rendah, serta mengungguli metode prapemrosesan tradisional (Hapsari, 2024).

Upaya peningkatan performa prediksi pada sektor farmasi juga terus berlanjut melalui pengembangan model hibrida. Salah satu pendekatan terbaru mengintegrasikan *Graph Convolutional Network* (GCN) dengan LSTM (KG-GCN-

LSTM) untuk memprediksi permintaan obat berdasarkan grafik pengetahuan relasional antara obat dan gejala klinis, yang terbukti secara signifikan meningkatkan akurasi prediksi dibandingkan model basis (Chen et al., 2025).

Secara keseluruhan, tinjauan terhadap berbagai literatur tersebut menunjukkan bahwa arsitektur LSTM dan variannya sangat andal dalam memproses data deret waktu penjualan obat. Meskipun demikian, mayoritas penelitian yang ada saat ini, khususnya di Indonesia, masih berfokus murni pada operasional manajemen persediaan dan peramalan permintaan stok obat. Pemanfaatan data penjualan dari distributor secara langsung sebagai parameter untuk memprediksi tren atau kecenderungan penyakit di suatu wilayah masih sangat terbatas. Oleh karena itu, dalam tabel 2.1 penelitian ini berupaya mengisi celah (*research gap*) tersebut dengan menggunakan metode LSTM tidak sekadar untuk mengamati tren penjualan, melainkan untuk mengekstraksi data tersebut menjadi indikator prediksi kecenderungan penyakit.

Tabel 2. 1 Tabel Penelitian Terdahulu

Peneliti (Tahun)	Fokus Penelitian	Data & Lokasi	Metode	Hasil Utama & Research Gap
Abbasi et al. (2023)	Sistem surveilans farmasi terintegrasi	Studi literatur sistem multinasional	Analisis kualitatif	Hasil: Data distributor valid untuk surveilans. Gap: Belum menerapkan model prediktif LSTM di Indonesia.
Kurniawati et al. (2023)	Prediksi penjualan obat di distributor	Data distributor obat di Indonesia	LSTM dan ARIMA	Hasil: RMSE LSTM lebih rendah dari ARIMA. Gap: Fokus pada stok obat, belum mengaitkan dengan tren penyakit.
Mbonyinshuti et al. (2024)	Peramalan permintaan rantai pasok	Data kesehatan di Rwanda, Afrika	LSTM dan ARIMA	Hasil: LSTM unggul menangkap fluktuasi permintaan.

				Gap: Konteks studi di luar negeri dan fokus operasional.
Hapsari (2024)	Prediksi permintaan dengan klasterisasi	Data industri farmasi Denpasar (60 bulan)	K-Means dan Bi-LSTM	Hasil: Bi-LSTM per klaster memberikan MAE/MSE rendah. Gap: Tujuan riset murni peramalan stok, bukan dominasi penyakit.
Chen et al. (2025)	Prediksi permintaan berbasis gejala	Data transaksi distributor obat	KG-GCN-LSTM	Hasil: Integrasi graf pengetahuan meningkatkan akurasi. Gap: Model belum diaplikasikan spesifik pada distributor lokal.

2.2 Landasan Teori

2.2.1 Prediksi Penyakit

Prediksi penyakit merupakan proses memperkirakan kemungkinan munculnya kejadian penyakit di masa mendatang berdasarkan pola historis yang tercatat dalam data kesehatan. Prediksi ini digunakan untuk membantu strategi pencegahan dan kesiapsiagaan sistem kesehatan. Pemodelan prediksi berbasis data kesehatan memungkinkan identifikasi awal terhadap pola kejadian penyakit, sehingga sistem kesehatan dapat mendeteksi kemungkinan terjadinya wabah lebih awal dan pengambilan keputusan preventif klinis dapat dilakukan secara tepat sasaran (Rajkomar et al., 2019; Property, 2015).

2.2.2 Data Penjualan Obat di Distributor sebagai Indikator Penyakit

Data transaksi obat di distributor dapat menggambarkan kondisi kesehatan masyarakat karena pembelian obat umumnya muncul sebagai respons langsung terhadap gejala penyakit. Pola penjualan obat bebas di jaringan distributor dapat digunakan sebagai sistem deteksi dini terhadap wabah penyakit, mengingat kenaikan pembelian obat tertentu sering kali mendahului

laporan klinis resmi dari rumah sakit. Oleh karena itu, pola pembelian obat dapat dijadikan sinyal atau indikator awal untuk memantau penyakit yang sedang mengalami tren peningkatan di wilayah tertentu (Wulandari et al., 2022).

2.2.3 Big Data dalam Kesehatan

Big data merupakan kumpulan data berskala besar dan kompleks yang memerlukan metode komputasi lanjutan untuk dikelola dan diolah. Di bidang kesehatan, integrasi *big data* memungkinkan pelaksanaan analisis prediktif yang krusial untuk deteksi penyakit dan perencanaan kesehatan masyarakat. Pemanfaatan *big data* memberikan peluang bagi sistem layanan kesehatan untuk memperoleh wawasan klinis dan epidemiologis melalui analitik berskala besar, yang pada akhirnya membuka potensi prediksi penyakit yang jauh lebih presisi dibandingkan metode konvensional (Dash et al., 2019; Wardhani et al., 2019).

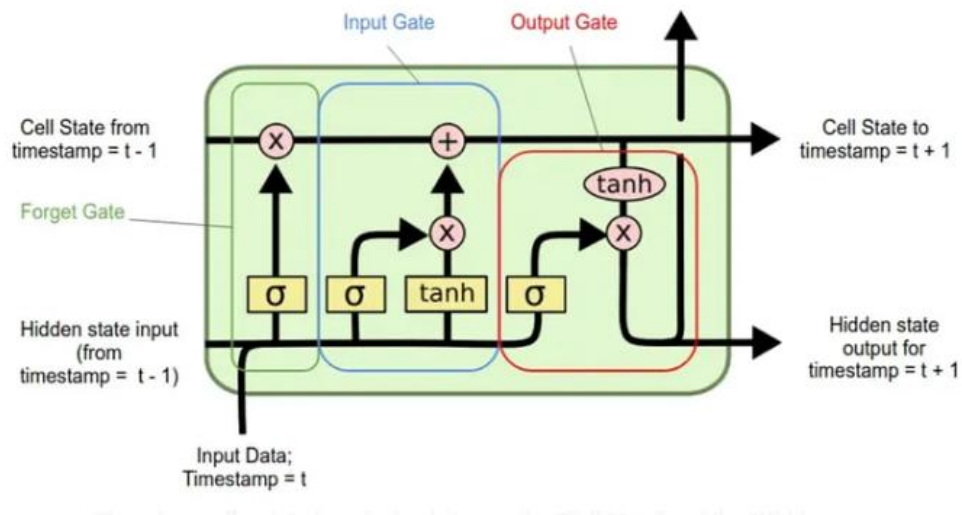
2.2.4 Machine Learning untuk Prediksi Penyakit

Machine learning merupakan salah satu cabang dari kecerdasan buatan yang berfokus pada pengembangan algoritma untuk mengidentifikasi pola tersembunyi dari suatu *dataset* dan menggunakannya untuk menghasilkan prediksi atau keputusan. Algoritma *machine learning*, termasuk jaringan saraf tiruan (*neural networks*) dan variannya, bekerja dengan mengenali kompleksitas pola data medis atau transaksi. Pendekatan komputasi ini telah terbukti mampu meningkatkan akurasi prediksi tren penyakit secara signifikan dibandingkan dengan metode statistik tradisional (Chen et al., 2021).

2.2.5 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) adalah arsitektur khusus dari *Recurrent Neural Network* (RNN) yang dirancang untuk mengatasi masalah *vanishing gradient* pada RNN tradisional. LSTM memiliki kemampuan untuk mempelajari dan menyimpan informasi dalam jangka panjang, menjadikannya sangat andal untuk memproses data berurutan (*sequential data*) seperti deret waktu (*time-series*). Algoritma ini sangat cocok diterapkan pada data historis

penjualan obat karena kemampuannya dalam memahami pola fluktuasi historis jangka panjang maupun jangka pendek (Siami-Namini et al., 2019; Mendoza-Silva et al., 2019).



Gambar 2. 1 Arsitektur LSTM

2.2.6 Sistem Prediksi Penyakit Berbasis LSTM pada Data Distributor

Integrasi arsitektur LSTM dengan data deret waktu penjualan obat bertujuan untuk memprediksi probabilitas munculnya penyakit berdasarkan pola pembelian secara otomatis. Penerapan model LSTM pada *dataset* transaksi penjualan distributor tidak hanya mampu memproyeksikan estimasi kebutuhan farmasi di masa depan, tetapi juga dapat mendeteksi kecenderungan penyakit dari waktu ke waktu. Pendekatan komputasi ini memberikan gambaran dinamika penyebaran penyakit dan fluktuasi kebutuhan obat yang lebih akurat, sehingga dapat diandalkan sebagai alat bantu dalam perencanaan intervensi layanan kesehatan (Fourkiotis & Tsadiras, 2024).

2.2.7 Metrik Evaluasi Model Prediksi

Untuk mengukur seberapa baik performa model klasifikasi LSTM dalam menebak kategori penyakit yang dominan, diperlukan perhitungan metrik tingkat ketepatan prediksi. Mengingat data logistik farmasi umumnya memiliki distribusi kelas yang tidak seimbang (*imbalanced dataset*), evaluasi

performa tidak dapat bergantung pada satu metrik tunggal. Oleh karena itu, penelitian ini menggunakan instrumen evaluasi *Confusion Matrix* beserta metrik turunan yang mencakup Akurasi (*Accuracy*), Presisi (*Precision*), Sensitivitas (*Recall*), dan *F1-Score*.

Dasar dari seluruh perhitungan evaluasi klasifikasi bertumpu pada *Confusion Matrix* (Matriks Kebingungan), yang merepresentasikan perbandingan hasil prediksi sistem dengan data aktual melalui empat parameter dasar: a. **True Positive (TP)**: Jumlah data positif yang diprediksi benar sebagai positif oleh model. b. **True Negative (TN)**: Jumlah data negatif yang diprediksi benar sebagai negatif oleh model. c. **False Positive (FP)**: Jumlah data negatif yang secara keliru diprediksi sebagai positif oleh model. d. **False Negative (FN)**: Jumlah data positif yang secara keliru diprediksi sebagai negatif oleh model.

Berdasarkan keempat parameter matriks tersebut, penjabaran masing-masing metrik evaluasi dirumuskan sebagai berikut:

2.2.7.1 Akurasi (*Accuracy*)

Akurasi adalah rasio antara jumlah prediksi yang benar (kategori penyakit diprediksi dengan tepat sesuai aktual) berbanding dengan keseluruhan total data uji. Metrik ini memberikan gambaran umum tentang ketepatan klasifikasi model. Secara matematis, akurasi dirumuskan sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

Semakin tinggi nilai persentase akurasi yang dihasilkan, semakin andal model tersebut dalam mengklasifikasikan tren penyakit yang akan mendominasi berdasarkan data historis yang ada (Chen et al., 2025).

2.2.7.2 Presisi (*Precision*)

Presisi mengukur tingkat ketepatan prediksi positif yang dihasilkan oleh model. Metrik ini mendeskripsikan rasio antara tebakan benar

(penyakit dominan) dibandingkan dengan keseluruhan prediksi positif yang dikeluarkan oleh sistem. Formula matematisnya adalah:

$$Precision = \frac{TP}{TP + FP}$$

2.2.7.3 Sensitivitas (*Recall*)

Sensitivitas atau *Recall* mengukur kemampuan model dalam mengidentifikasi seluruh kasus positif pada data aktual. Metrik ini mendeskripsikan rasio antara prediksi positif yang benar dibandingkan dengan total keseluruhan kejadian nyata di lapangan. Formula matematisnya adalah:

$$Recall = \frac{TP}{TP + FN}$$

2.2.7.4 *F1-Score*

F1-Score merupakan nilai rata-rata harmonik (*harmonic mean*) dari metrik Presisi dan Sensitivitas. Metrik ini memberikan evaluasi komprehensif yang lebih objektif dibandingkan Akurasi, terutama saat terjadi ketimpangan jumlah sampel antar kategori sindrom penyakit. Model dianggap memiliki stabilitas prediksi yang baik apabila nilai *F1-Score* menunjukkan persentase yang tinggi. Formula matematisnya adalah:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

2.2.8 eXtreme Gradient Boosting (XGBoost)

eXtreme Gradient Boosting (XGBoost) adalah algoritma pembelajaran mesin berbasis *ensemble* pohon keputusan (*decision tree*) yang menggunakan prinsip *gradient boosting*. Berbeda dengan jaringan saraf tiruan yang memproses aliran data secara sekuensial, XGBoost bekerja dengan cara membangun banyak pohon keputusan secara berurutan, di mana setiap pohon

baru secara iteratif dirancang untuk mengoreksi galat (*error*) atau residu klasifikasi dari pohon-pohon sebelumnya.

Algoritma ini diakui secara luas sebagai standar utama (*state-of-the-art*) dalam menangani pemodelan data berformat tabular yang terstruktur. Keunggulan komputasional utama dari XGBoost terletak pada kemampuannya untuk memetakan hubungan *non-linear* yang kompleks melalui penetapan batas ambang (*threshold*) fitur secara spesifik dan independen. Selain itu, algoritma ini dilengkapi dengan mekanisme regularisasi matematis bawaan (L1 dan L2) untuk menekan risiko penghafalan data (*overfitting*). Dalam konteks pemrosesan data logistik atau fluktuasi angka absolut, pendekatan batas ambang pada XGBoost membuatnya sangat responsif dalam mendeteksi lonjakan anomali seketika tanpa terpengaruh oleh kelembaman (*lag*) dari riwayat waktu masa lalu.

Dalam literatur riset prediktif, XGBoost secara luas diakui sebagai *baseline* (model pembanding dasar) yang paling mumpuni untuk menguji keandalan sebuah arsitektur *Deep Learning*. Pemilihan XGBoost sebagai algoritma komparasi tunggal pada penelitian ini didasarkan pada perbedaan mendasar secara arsitektural. LSTM mengandalkan struktur jaringan saraf sekuensial yang sangat baik dalam "mengingat" pola siklus jangka panjang, namun rentan mengalami kelembaman (*lagging*) terhadap anomali sesaat. Sebaliknya, XGBoost merupakan representasi terbaik dari algoritma *non-deep-learning* berbasis pohon (*tree-based*) yang memproses data secara parsial dan tabular. Mengomparasikan kedua pendekatan komputasi yang saling bertolak belakang ini (sekuensial vs tabular) akan membuktikan secara objektif metode mana yang paling relevan dan presisi untuk menangkap dinamika wabah dari sebuah rekam data logistik farmasi.

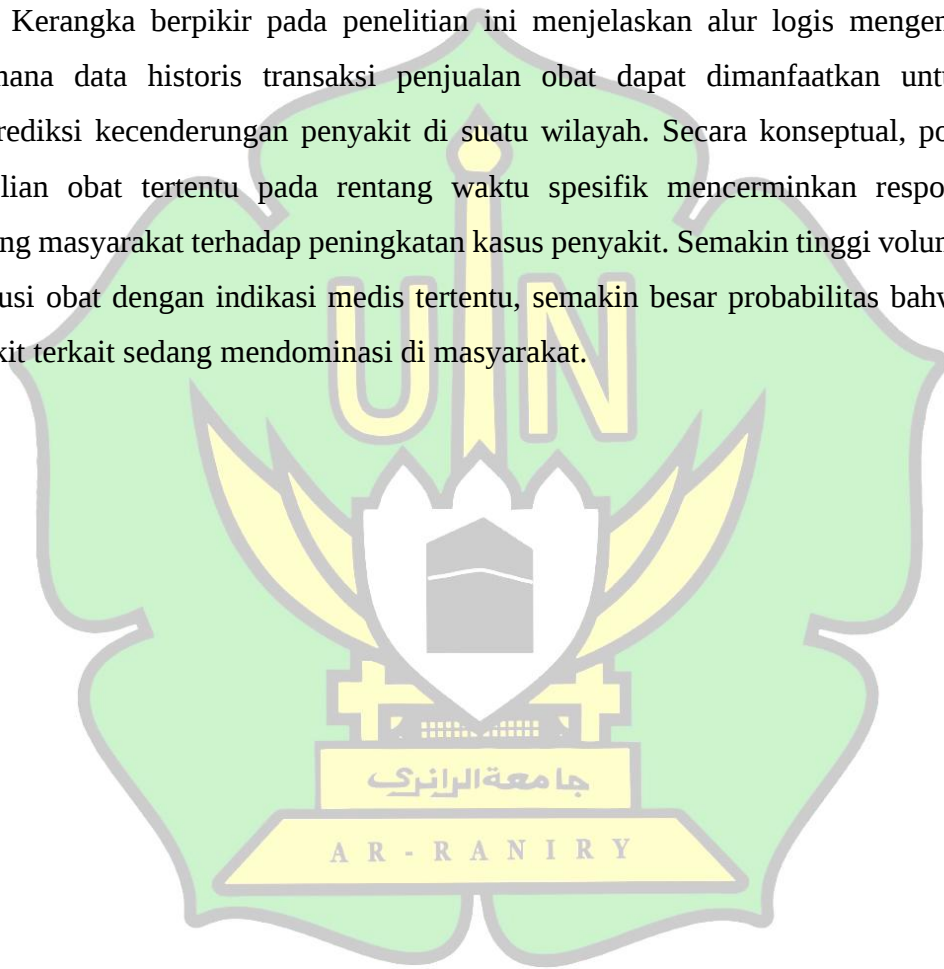
2.2.9 Lingkungan Pengembangan Model (Google Colaboratory)

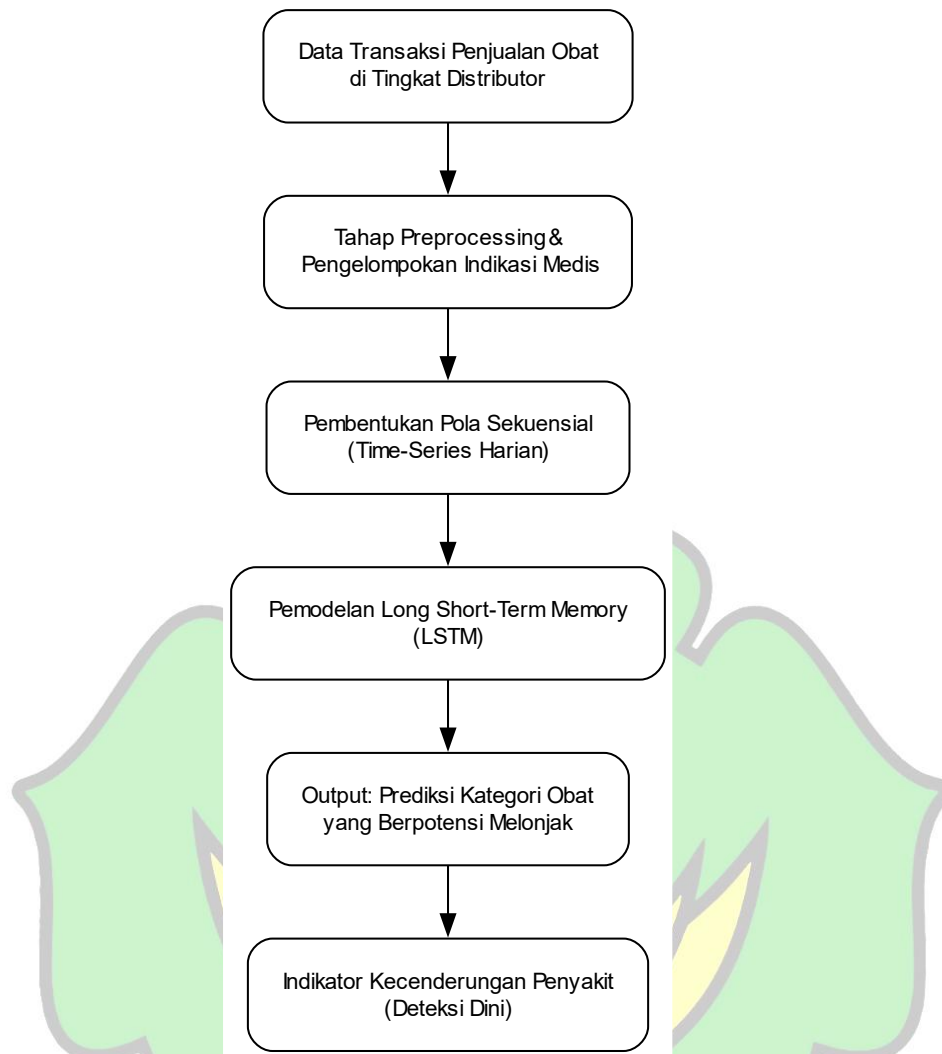
Google Colaboratory (Colab) merupakan platform lingkungan pengembangan berbasis *cloud* yang disediakan oleh Google, yang memungkinkan eksekusi kode bahasa pemrograman Python secara langsung melalui *browser*. Platform ini menyediakan akses komputasi berbasis *Graphics*

Processing Unit (GPU) yang krusial untuk mempercepat proses pelatihan (*training*) model *deep learning* berukuran besar seperti LSTM, tanpa membebani perangkat keras lokal. Pemodelan ini juga didukung oleh ekosistem *library* Python, khususnya TensorFlow/Keras, yang menyediakan modul *high-level* untuk membangun dan melatih arsitektur jaringan saraf tiruan secara efisien.

2.3 Kerangka Berpikir

Kerangka berpikir pada penelitian ini menjelaskan alur logis mengenai bagaimana data historis transaksi penjualan obat dapat dimanfaatkan untuk memprediksi kecenderungan penyakit di suatu wilayah. Secara konseptual, pola pembelian obat tertentu pada rentang waktu spesifik mencerminkan respons langsung masyarakat terhadap peningkatan kasus penyakit. Semakin tinggi volume distribusi obat dengan indikasi medis tertentu, semakin besar probabilitas bahwa penyakit terkait sedang mendominasi di masyarakat.





Gambar 2. 2 Alur logika rangka pemikiran

Berdasarkan alur logika tersebut, penelitian ini memproses data penjualan obat yang telah melalui tahap pembersihan (*data cleaning*) dan pengelompokan berdasarkan indikasi penyakit. Data tersebut diagregasi berdasarkan periode waktu (harian) untuk membentuk pola sekuensial deret waktu (*time-series*). Pola historis ini selanjutnya diproses menggunakan arsitektur *Long Short-Term Memory* (LSTM) untuk mempelajari hubungan keterulangan dan fluktuasi dari waktu ke waktu. *Output* dari pemodelan ini berupa prediksi kategori obat yang berpotensi mengalami lonjakan permintaan pada periode selanjutnya, yang secara paralel berfungsi sebagai indikator awal (deteksi dini) terhadap potensi penyebaran penyakit di wilayah cakupan distributor. Hasil prediksi ini divisualisasikan dan dievaluasi tingkat kesalahannya menggunakan metrik statistik.

2.4 Hipotesis

Penelitian ini merupakan penelitian kuantitatif komputasional yang berfokus pada pemodelan prediktif. Hipotesis disusun untuk menguji kemampuan dan validitas model jaringan saraf tiruan dalam melakukan proyeksi berdasarkan data historis. Mengacu pada rumusan masalah yang telah ditetapkan, hipotesis yang diajukan dalam penelitian ini adalah sebagai berikut:

1. **Hipotesis Nol (H_0):** Model arsitektur *Long Short-Term Memory* (LSTM) tidak mampu memprediksi kecenderungan penyakit berdasarkan data historis penjualan obat di tingkat distributor secara valid (menghasilkan metrik evaluasi Akurasi yang rendah dan tidak merepresentasikan pola data sesungguhnya).
2. **Hipotesis Alternatif (H_1):** Model arsitektur *Long Short-Term Memory* (LSTM) mampu memprediksi kecenderungan penyakit berdasarkan data historis penjualan obat di tingkat distributor secara valid dan signifikan, yang dibuktikan dengan metrik evaluasi Akurasi (*Accuracy*) yang tinggi.



BAB III

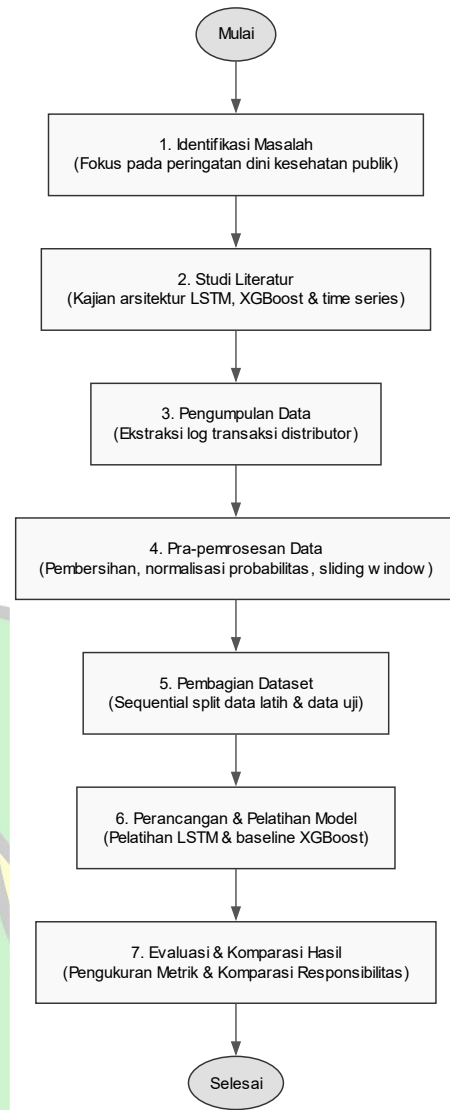
METODOLOGI PENELITIAN

3.1 Jenis Penelitian

Penelitian ini termasuk dalam kategori **penelitian kuantitatif prediktif** yaitu penelitian yang memanfaatkan data numerik dan model matematis untuk memprediksi suatu fenomena berdasarkan pola historis. Pendekatan ini digunakan untuk membangun model perkiraan peningkatan penyakit berdasarkan pola penjualan obat pada periode tertentu di distributor. Model prediksi yang dibangun diharapkan mampu memberikan gambaran kecenderungan penyakit yang muncul dalam kurun waktu mendatang.

3.2 Alur Penelitian

Penelitian ini menggunakan pendekatan kuantitatif prediktif dengan pemodelan peramalan deret waktu (*time series forecasting*). Data transaksi distribusi obat diolah dan diurutkan secara kronologis untuk mengidentifikasi pola frekuensi dan lonjakan permintaan. Pendekatan ini dilandasi oleh asumsi bahwa setiap fluktuasi signifikan pada volume distribusi obat-obatan tertentu memiliki korelasi langsung dengan dinamika dan perubahan tren gejala penyakit akut di masyarakat. Melalui desain komputasional ini, alur hubungan historis antarvariabel dipelajari secara algoritmik guna menghasilkan model prediksi yang terukur, dapat diuji validitasnya, serta direplikasi secara ilmiah.



Gambar 3. 1 Skema Alur Penelitian

Berdasarkan Gambar 3.1, rangkaian operasional dalam alur penelitian ini disusun secara sistematis melalui tahapan-tahapan berikut:

1. **Identifikasi Masalah:** Tahap ini bertujuan untuk merumuskan urgensi penelitian berdasarkan kesenjangan antara kondisi faktual dan potensi ideal. Permasalahan utama yang diangkat adalah belum optimalnya pemanfaatan data rekam jejak distribusi obat di tingkat distributor lokal. Selama ini, data tersebut murni digunakan sebagai instrumen pencatatan administratif dan manajemen inventaris. Padahal, dengan pendekatan analitik lanjutan, pergerakan data farmasi tersebut memiliki potensi strategis sebagai

indikator peringatan dini (*early warning system*) terhadap potensi lonjakan kasus penyakit di masyarakat sebelum data klinis resmi dirilis.

2. **Studi Literatur:** Kajian pustaka dilakukan secara komprehensif untuk membangun landasan teori komputasi dan menemukan celah penelitian (*research gap*). Fokus telaah diarahkan pada publikasi ilmiah terkini yang membahas implementasi pemelajaran mesin pada deret waktu, khususnya arsitektur *Long Short-Term Memory* (LSTM) dan algoritma komparasi *eXtreme Gradient Boosting* (XGBoost) dalam domain informatika kesehatan. Kajian literatur yang sistematis ini esensial untuk membandingkan parameter model, memahami keterbatasan metode terdahulu, dan menjustifikasi pemilihan arsitektur model yang akan dibangun.
3. **Pengumpulan Data:** Fase pengumpulan data dilakukan melalui prosedur dokumentasi dan ekstraksi dari sistem basis data distributor farmasi. Data yang ditarik merupakan data sekunder berwujud log transaksi historis. Variabel utama yang diekstrak mencakup stempel waktu (*timestamp*), identitas produk obat, indikasi medis atau kategori obat, serta kuantitas distribusi. Keutuhan dan keakuratan ekstraksi pada tahap ini sangat menentukan kualitas *dataset* yang akan menjadi input bagi algoritma pembelajaran mesin.
4. **Pembersihan dan Pra-pemrosesan Data (*Data Preprocessing*):** Data mentah hasil ekstraksi umumnya masih memuat *noise*, inkonsistensi, atau format yang tidak standar. Oleh karena itu, dilakukan pra-pemrosesan yang meliputi penanganan data kosong (*handling missing values*), penghapusan data duplikat, dan agregasi data ke dalam interval waktu yang seragam. Selanjutnya, dilakukan transformasi nilai masukan menggunakan penskalaan proporsi probabilitas matematis dan pembentukan struktur sekuensial (*sliding window*) agar sesuai dengan syarat masukan (*input shape*) arsitektur LSTM.
5. **Pembagian Dataset (*Train-Test Split*):** Untuk mengevaluasi kemampuan generalisasi model, *dataset* yang telah diproses dibagi menjadi

kompartemen utama: data latih (*training set*) dan data uji (*testing set*). Berbeda dengan pembagian data pada umumnya yang dilakukan secara acak, pembagian pada penelitian *time series* ini dilakukan secara berurutan (*sequential split*) untuk mempertahankan integritas kronologis dari pola deret waktu. Proses ini krusial untuk mencegah model sekadar menghafal data (*overfitting*).

6. **Perancangan dan Pelatihan Model Komparatif:** Tahap inti komputasi berada pada perancangan arsitektur prediksi algoritmik. *Deep learning* LSTM dipilih karena memiliki mekanisme memori internal (*gates*) yang spesifik dirancang untuk mengatasi masalah degradasi gradien (*vanishing gradient*) pada data sekuensial jangka panjang. Sebagai pembanding validasi, algoritma XGBoost juga dirancang untuk menangkap anomali tabular menggunakan prinsip *gradient boosting*. Kedua model ini kemudian dilatih secara terpisah menggunakan data latih untuk mengenali pola *non-linear* dari fluktuasi permintaan obat.
7. **Evaluasi dan Komparasi Hasil:** Tahap akhir adalah pengujian performa kedua model menggunakan data uji. Tingkat ketepatan prediksi diukur secara matematis menggunakan metrik klasifikasi multikelas yang terdiri dari *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Luaran evaluasi dari model LSTM dan XGBoost kemudian dikomparasikan secara langsung (*head-to-head*) untuk menginterpretasi algoritma mana yang paling responsif dan memiliki galat terkecil dalam mendeteksi transisi penyakit di lapangan.

3.3 Populasi dan Sampel

3.3.1 Populasi Penelitian

Populasi dalam penelitian ini adalah keseluruhan rekam data sekunder transaksi distribusi obat yang tercatat pada sistem informasi manajemen distributor farmasi di wilayah Banda Aceh. Populasi ini tidak merujuk pada subjek manusia (pasien), melainkan pada keseluruhan entri log transaksi mentah berjumlah 94.041 baris yang mencakup atribut waktu, jenis obat, dan volume penjualan yang merepresentasikan dinamika permintaan di masyarakat.

3.3.2 Sampel Penelitian

Sampel penelitian diambil menggunakan teknik *non-probability sampling* dengan pendekatan *purposive sampling*. Kriteria inklusi sampel adalah log transaksi obat yang memiliki indikasi medis terhadap penyakit akut tertentu (enam kategori sindrom), memiliki atribut tanggal yang lengkap, dan terdistribusi selama periode observasi (Januari 2023 hingga Desember 2024). Pemilihan sampel terarah ini dilakukan secara algoritmik untuk memastikan model terhindar dari *noise* obat-obatan yang tidak relevan dengan tren penyakit.

Mengingat pemodelan deret waktu sangat bergantung pada dependensi berurutan, sampel yang memenuhi kriteria inklusi tersebut tidak diambil secara acak, melainkan digunakan secara keseluruhan (*total sampling* dari *subset* yang difilter). Pendekatan ini merupakan syarat optimal untuk mempertahankan karakteristik temporal dalam melatih arsitektur jaringan saraf. Melalui proses penyaringan data mentah tersebut, jumlah sampel log transaksi yang secara empiris digunakan dalam pelatihan dan pengujian arsitektur komputasi ini adalah data-data yang murni merepresentasikan keenam sindrom penyakit..

3.4 Teknik Pengumpulan Data

3.4.1 Sumber Data

Sumber data yang digunakan dalam penelitian ini murni merupakan data sekunder berupa rekam transaksi distribusi obat historis dari PT. Kimia Farma Trading & Distribution cabang Banda Aceh. Keseluruhan data populasi yang berhasil diekstrak berjumlah 94.041 baris rekaman transaksi. Seluruh berkas data mentah (raw data) berformat spreadsheet tersebut disimpan dan didokumentasikan secara terpusat pada repositori cloud Google Drive untuk memfasilitasi proses impor ke lingkungan komputasi Google Colab, yang dapat diakses melalui tautan: <https://docs.google.com/spread-sheets/d/1b6ZNzlyHkvNk6RaEtEl8nZi2vVYZgT8/edit?usp=sharing&ouid=101195567669182579794&rtpof=true&sd=true>

3.4.2 Prosedur Ekstraksi dan Dokumentasi

Pengumpulan data dieksekusi melalui prosedur penarikan data (*data export*) secara langsung dari antarmuka sistem manajemen basis data internal distributor. Berkas hasil ekstraksi tersebut masih berwujud *spreadsheet* mentah yang belum melalui proses pembersihan (*data cleaning*), sehingga masih memuat format penamaan kolom bawaan mesin (*dirty headers*) serta atribut logistik yang tidak berkaitan langsung dengan analisis kecenderungan penyakit.

Secara spesifik, log transaksi mentah hasil ekstraksi memuat sembilan atribut kolom utama yang rinciannya diuraikan pada Tabel 3.1 berikut:

Tabel 3. 1 Deskripsi Atribut Kolom Dataset Mentah Distributor

No	Header Tabel	Penjelasan
1	No	Nomor urut baris data (indeks baris log transaksi).
2	Posting Date	Stempel waktu (<i>timestamp</i>) atau tanggal dibukukannya transaksi distribusi obat pada sistem kasir/gudang.
3	Material	Kode unik logistik atau identitas numerik (ID) dari produk obat dalam sistem basis data distributor (SKU).
4	Material description	Nama atau deskripsi produk obat yang didistribusikan, yang memuat informasi merek, jenis sediaan, serta dosis.
5	Plant	Kode unit fasilitas gudang atau lokasi cabang distributor farmasi (misalnya representasi kode cabang Banda Aceh).
6	Qty in unit of entry	Jumlah kuantitas produk yang dikeluarkan berdasarkan satuan logistik besar. Pada sistem ERP, transaksi keluar (<i>goods issue</i>) umumnya dicatat dengan nilai negatif.
7	Unit of Entry	Satuan ukur kemasan logistik komersial yang digunakan saat input transaksi (misalnya: DUS, BOX, atau KARTON).
8	Base Unit of Measure	Satuan ukuran dasar atau satuan eceran terkecil dari obat yang didistribusikan (misalnya: TAB untuk tablet, TUB untuk <i>tube</i> , AMP untuk ampul).
9	Quantity	Total agregat kuantitas unit obat yang didistribusikan (dihitung dalam nilai satuan dasar terkecil). Nilai ini yang akan dikonversi menjadi proporsi dalam pemodelan algoritma.

Berkas data dengan struktur bawaan operasional ini kemudian didokumentasikan sebagaimana aslinya. Seleksi, pembuangan kolom yang tidak relevan (seperti Plant, Material, atau Unit of Entry), dan restrukturisasi kolom baru akan dieksekusi secara programatis pada tahap pra-pemrosesan menggunakan bahasa pemrograman Python.

3.5 Teknik Analisis Data

Analisis data merupakan tahapan komputasi inti dalam penelitian ini yang bertujuan untuk mengolah data historis transaksi distribusi obat menjadi model prediksi kecenderungan penyakit. Rangkaian analisis dilakukan secara sistematis mulai dari tahap pra-pemrosesan, transformasi deret waktu, pembagian *dataset*, pelatihan arsitektur *Long Short-Term Memory* (LSTM), hingga evaluasi dan visualisasi. Setiap tahapan dieksekusi untuk memastikan model yang dibangun tidak hanya valid secara matematis, tetapi juga bebas dari bias sehingga dapat diaplikasikan secara praktis sebagai alat bantu pengambilan keputusan strategis di tingkat distributor.

3.5.1 Pra-Pemrosesan Data (*Data Preprocessing*)

Tahap ini merupakan fondasi komputasi untuk mengubah log transaksi mentah menjadi matriks fitur yang terstruktur dan siap dipelajari oleh model jaringan saraf tiruan. Proses ini dimulai dengan kategorisasi data, di mana deskripsi material obat dikelompokkan ke dalam enam kelas sindrom farmakologis menggunakan teknik pencocokan kata kunci (*keyword matching*). Selanjutnya, data diagregasi berdasarkan interval waktu temporal (mingguan atau bulanan). Penanganan nilai kosong (*missing values*) pada interval yang tidak memiliki rekam transaksi diselesaikan menggunakan metode substitusi nilai nol (*zero imputation*).

Berbeda dengan normalisasi skalar konvensional, transformasi nilai masukan pada penelitian ini dilakukan dengan mengonversi kuantitas volume absolut menjadi matriks proporsi probabilitas. Melalui pembagian nilai kategori dengan total volume per baris waktu, data direpresentasikan dalam rentang nilai 0 hingga 1. Pendekatan ini memastikan arsitektur jaringan tidak terganggu oleh

lonjakan volume logistik yang acak, melainkan terkalibrasi murni pada pergeseran dominasi tren penyakit. Sebagai pelengkap, ekstraksi sinyal waktu dilakukan dengan menormalisasi nilai indeks bulan kalender melalui pembagian skalar (dibagi 12), guna membantu algoritma mengenali siklus musiman secara stabil.

3.5.2 Transformasi Data Time Series

Data yang telah dibersihkan kemudian ditransformasi ke dalam format sekuensial dengan indeks waktu mingguan. Proses ini mencakup agregasi volume distribusi obat yang dikelompokkan berdasarkan kategori sindrom farmakologis dan dinormalisasi menjadi bentuk persentase proporsi. Untuk memenuhi syarat masukan model LSTM, data dibentuk menggunakan algoritma *sliding window* (4 minggu) dengan pendekatan multivariat untuk menciptakan variabel fitur *lag*, yang memuat probabilitas historis setiap sindrom beserta representasi waktunya..

3.5.3 Pembagian Dataset (*Training–Testing Split*)

Dataset yang telah berformat deret waktu dipartisi secara sekuensial (berurutan) berdasarkan kalender operasional untuk menjaga integritas temporal dan mencegah terjadinya kebocoran data (*data leakage*). Berbeda dengan pemodelan pada data statis yang membagi sampel secara acak, pembagian pada penelitian deret waktu ini membagi kompartemen data ke dalam tiga fungsi spesifik: data latih (*training set*), data validasi (*validation set*), dan data uji (*testing set*).

Secara operasional, seluruh matriks log transaksi dari rentang waktu tahun 2023 dialokasikan sebagai ruang pembelajaran model. Di dalam kompartemen tahun 2023 tersebut, sistem menerapkan pemisahan temporal dengan proporsi 80:20. Sebesar 80% data pada periode awal tahun digunakan murni sebagai data latih untuk mengoptimasi bobot jaringan saraf, sedangkan 20% sisa data pada periode akhir tahun digunakan sebagai data validasi internal. Penggunaan data validasi ini krusial untuk memantau pergerakan konvergensi

dan memastikan model tidak sekadar menghafal data (*overfitting*) selama siklus pelatihan (*epoch*) berlangsung.

Sementara itu, matriks data historis yang merangkum distribusi sepanjang tahun 2024 diisolasi secara mutlak sejak awal pra-pemrosesan. Keseluruhan data 2024 ini ditetapkan sebagai data uji akhir (*testing set*) yang benar-benar asing dan belum pernah dipelajari oleh algoritma sebelumnya. Pemisahan kompartemen lintas tahun ini memastikan bahwa simulasi evaluasi prediksi nantinya benar-benar mencerminkan kondisi inferensi dunia nyata terhadap data di masa depan.

3.5.4 Perancangan dan Pelatihan Model LSTM

Tahap komputasi berlanjut pada perancangan arsitektur jaringan LSTM. Algoritma ini diimplementasikan karena desain sel memorinya mumpuni dalam menangkap dependensi jangka panjang pada data transaksi obat yang fluktuatif. Parameter jaringan seperti jumlah *hidden layer*, *neuron*, *learning rate*, dan ukuran *batch* dikonfigurasi secara manual. Model kemudian dilatih menggunakan algoritma optimasi BPTT melalui lingkungan eksekusi berbasis *cloud* dengan akselerasi GPU. Pembaruan bobot dieksekusi iteratif berdasarkan fungsi kerugian *Sparse Categorical Crossentropy* hingga mencapai konvergensi optimal.

3.5.5 Evaluasi Model Prediksi

Setelah proses iterasi pelatihan selesai, model diuji menggunakan data pengujian dari tahun yang diisolasi. Tingkat keandalan prediksi diukur secara kuantitatif menggunakan metrik evaluasi utama klasifikasi, yaitu Akurasi. Penggunaan metrik ini memberikan gambaran objektif mengenai seberapa tepat persentase tebakan klasifikasi model dibandingkan dengan kondisi empiris sebenarnya di lapangan.

Selain mengukur instrumen metrik internal dari model tunggal, tahap evaluasi pada penelitian ini juga diperluas menggunakan metode analisis komparatif (*comparative analysis*). Hasil inferensi dari arsitektur sekuensial *Long Short-Term Memory* (LSTM) akan dipertandingkan secara langsung

(*head-to-head*) dengan luaran klasifikasi dari algoritma *eXtreme Gradient Boosting* (XGBoost). Pemilihan XGBoost sebagai satu-satunya model pembandingan (*baseline*) didasarkan pada pertimbangan akademis bahwa algoritma tersebut merupakan algoritma berbasis pohon (*tree-based*) terkuat untuk data tabular. Komparasi ini dirancang secara spesifik untuk menguji hipotesis algoritmik: apakah arsitektur memori berurutan (LSTM) benar-benar memberikan keunggulan performa pada data *time-series*, atau justru kalah dari algoritma *non-deep-learning* (XGBoost) yang mengabaikan sekuensi masa lalu namun sangat sensitif terhadap batas ambang (*threshold*) lonjakan mendadak di lapangan. Melalui komparasi ini, kelemahan operasional dari masing-masing algoritma dapat diidentifikasi secara menyeluruh.

3.5.6 Interpretasi Hasil dan Visualisasi

Selain evaluasi numerik, keluaran dari pemodelan ini direpresentasikan secara visual untuk memudahkan analisis tren. Visualisasi diwujudkan dalam bentuk plot garis (*line plot*) yang membandingkan fluktuasi data kejadian aktual dengan pergerakan hasil prediksi algoritma. Representasi grafis ini dirancang untuk menerjemahkan deretan angka probabilitas menjadi instrumen interpretasi peringatan dini yang lebih intuitif.

3.6 Objek dan Subjek Penelitian

3.6.1 Objek Penelitian

Objek penelitian ini adalah pola data transaksi penjualan obat dan hasil prediksi kecenderungan penyakit. Data yang menjadi sasaran analisis meliputi atribut transaksi seperti *timestamp*, jenis produk obat, indikasi medis, dan volume unit yang terdistribusi. Objek data tersebut diperlakukan sebagai data sekuensial yang diolah menggunakan arsitektur LSTM untuk mempelajari dependensi jangka panjang. Penggunaan pola transaksi ini ditetapkan atas dasar logis bahwa fluktuasi permintaan berkorelasi langsung dengan lonjakan kasus klinis; ketika suatu penyakit menyebar, permintaan terhadap farmakologi terkait akan meningkat secara terukur.

3.6.2 Subjek Penelitian

Subjek dalam penelitian ini adalah distributor farmasi yang berlokasi di wilayah Banda Aceh (PT. Kimia Farma Trading & Distribution), yang bertindak sebagai entitas penyedia rekam data transaksi. Penelitian ini tidak menjadikan individu pasien secara langsung sebagai subjek, sehingga bersifat non-intervensi dan menjamin kerahasiaan. Pemilihan entitas logistik sebagai subjek didasarkan pada besarnya kapasitas institusi dalam mendistribusikan obat ke fasilitas kesehatan secara regional, sehingga agregasi datanya sangat memadai untuk simulasi prediksi.

3.6.3 Kriteria Subjek dan Objek Data

Agar *dataset* yang digunakan relevan dan valid untuk pemodelan algoritma jaringan saraf tiruan, penelitian ini menetapkan batasan kriteria ekstraksi data sebagai berikut:

1. Keceragaman Sumber: Data transaksi ditarik dari satu sistem basis data distributor yang sama untuk menjaga konsistensi format dan integritas pencatatan.
2. Struktur Temporal: Data memiliki kelengkapan atribut waktu yang berkesinambungan sehingga dapat diagregasi menjadi deret waktu (harian atau mingguan).
3. Klasifikasi Indikasi: Data mencakup informasi kategori terapeutik atau indikasi medis obat yang terdefinisi dengan jelas (misalnya: penurun demam, analgesik, atau obat pernapasan).
4. Anonimitas: Seluruh entri data murni memuat angka logistik dan tidak mengandung informasi identitas pribadi (*Personally Identifiable Information*), sehingga sepenuhnya mematuhi standar prosedur komputasi dan etika privasi data observasional..

3.7 Lokasi dan Waktu Penelitian

3.7.1 Lokasi Penelitian

Penelitian ini dilaksanakan dengan mengambil studi kasus pada PT. Kimia Farma Trading & Distribution yang berlokasi di Jl. Pemancar No. 11, Lamteumen Timur, Kota Banda Aceh. Pemilihan lokasi institusi ini didasarkan pada kapasitas distributor tersebut dalam menyuplai obat-obatan ke berbagai fasilitas kesehatan di wilayah Banda Aceh, serta ketersediaan sistem pencatatan transaksi yang terdokumentasi secara sistematis. Data dari lokasi ini dinilai sangat representatif untuk diekstraksi menjadi data runtun waktu (*time series*) guna menganalisis pola penjualan obat sebagai indikator kecenderungan penyakit.

3.7.2 Waktu Penelitian

Pelaksanaan penelitian ini, mulai dari tahap perencanaan, pengumpulan data, simulasi komputasi, hingga penyusunan laporan akhir, direncanakan berlangsung selama lima bulan, yaitu terhitung sejak bulan Januari 2026 hingga Mei 2026.

3.8 Perangkat yang Digunakan (Software dan Hardware)

Dalam pelaksanaan penelitian ini, beban komputasi utama untuk pemrosesan deret waktu dan pelatihan arsitektur *Long Short-Term Memory* (LSTM) dieksekusi pada lingkungan *cloud*. Oleh karena itu, spesifikasi diuraikan menjadi dua bagian: perangkat lunak yang mencakup ekosistem *library* Python, dan perangkat keras dengan spesifikasi minimal yang dibutuhkan untuk menjalankan klien antarmuka.

3.8.1 Perangkat Lunak (Software)

Perangkat lunak yang digunakan berfokus pada ekosistem pembelajaran mesin (*machine learning*) berbasis Python. Mengingat pengembangan dilakukan di Google Colaboratory, seluruh modul dan *library* yang diperlukan diuraikan pada Tabel 3.3 berikut:

Tabel 3. 2 Spesifikasi Perangkat Lunak

Kategori	Nama Software	Fungsi / Deskripsi
----------	---------------	--------------------

Lingkungan Pengembangan	Google Colaboratory (<i>Colab</i>)	Platform IDE berbasis <i>cloud</i> (Jupyter Notebook) untuk eksekusi kode tanpa membebani komputasi lokal.
Bahasa Pemrograman	Python 3.x	Bahasa pemrograman utama untuk pengolahan data dan pemodelan matematis.
Pemrosesan Data	Pandas, NumPy	Manipulasi struktur data tabular, agregasi waktu, dan normalisasi proporsi matematis.
Pemodelan Deep Learning	TensorFlow / Keras	Kerangka kerja (<i>framework</i>) utama untuk membangun, mengompilasi, dan melatih arsitektur sel memori LSTM.
Visualisasi Data	Matplotlib, Seaborn	Pembuatan representasi grafis berupa plot garis tren prediksi, konvergensi pelatihan, dan fluktuasi aktual.
Manajemen Data Mentah	Microsoft Excel	Pemeriksaan awal, verifikasi integritas kolom transaksi, dan peninjauan format data CSV/XLSX dari distributor.
Evaluasi Model	Sckit-Learn	Pustaka (<i>Library</i>) yang digunakan untuk melakukan ekstraksi <i>classification report</i> dan metrik performa menggunakan <i>confusion metrics</i>

3.8.2 Perangkat Keras (*Hardware*)

Karena proses pelatihan (*training*) jaringan saraf tiruan dilakukan secara *online* pada server Google, perangkat keras lokal hanya bertindak sebagai peramban klien (*browser client*). Tabel 3.4 menguraikan spesifikasi minimal (lokal) dan sumber daya *cloud* yang dialokasikan.

Tabel 3. 3 Spesifikasi Perangkat Keras Minimal

Komponen	Spesifikasi Minimal yang Digunakan
Prosesor (Lokal)	Setara Intel Core i3 / i5 atau AMD Ryzen 3 / 5
RAM (Lokal)	8 GB
Sistem Operasi	Linux ekuivalen, atau Windows 10/11 (mendukung peramban web modern)
Penyimpanan (Lokal)	256 GB SSD (untuk manajemen <i>dataset</i> lokal dan penyimpanan skrip <i>backup</i>)

Konektivitas Jaringan	Koneksi internet stabil (syarat mutlak untuk menjaga sesi Google Colab tidak terputus saat BPTT)
Penyimpanan (Cloud)	Google Drive (dimuat otomatis/ <i>mounted</i> ke Colab untuk direktori <i>dataset</i> dan ekspor <i>model weights</i>)
Komputasi GPU (Cloud)	Akselerator GPU yang disediakan oleh Google Colab (misalnya NVIDIA T4 Tensor Core) untuk mempercepat pelatihan LSTM

3.9 Variabel Penelitian

Dalam konteks pemodelan *machine learning* berbasis deret waktu (*time series*), variabel penelitian diklasifikasikan ke dalam fitur masukan (*input features*) dan target keluaran (*output target*), yang diuraikan sebagai berikut:

3.9.1 Variabel Independen (Variabel Bebas)

Variabel independen dalam penelitian ini adalah data historis distribusi obat pada langkah waktu sebelumnya (*lag features*). Variabel ini bertindak sebagai parameter masukan multivariat bagi arsitektur *Long Short-Term Memory* (LSTM) untuk dipelajari pola keterulangannya, yang meliputi:

1. Riwayat proporsi volume kuantitas obat dari enam kategori sindrom yang didistribusikan pada periode waktu $t - n$.
2. Nilai skalar indeks waktu (bulan) dari transaksi tersebut yang telah dinormalisasi.

3.9.2 Variabel Dependen (Variabel Terikat)

Variabel dependen dalam penelitian ini adalah kelas target prediktif yang dihasilkan oleh model pada periode waktu yang akan datang ($t + 1$). Variabel ini meliputi klasifikasi kategori kecenderungan penyakit yang paling dominan di masyarakat, yang direpresentasikan melalui identifikasi label indeks klasifikasi dari kelompok obat dengan probabilitas lonjakan tertinggi.

3.10 Instrumen Penelitian

Mengingat penelitian ini murni bersifat komputasional kuantitatif, instrumen yang digunakan untuk pengumpulan dan pengolahan data tidak berupa kuesioner, melainkan terdiri dari:

1. **Dataset Transaksi Distribusi:** Rekam data historis sekunder dari PT. Kimia Farma Trading & Distribution cabang Banda Aceh, yang bertindak sebagai instrumen observasi utama terhadap pola fluktuasi kebutuhan obat.
2. **Skrip Program (Source Code):** Susunan algoritma berbasis bahasa pemrograman Python yang memuat arsitektur pemodelan Long Short-Term Memory (LSTM) untuk klasifikasi deret waktu, yang berfungsi sebagai instrumen analitik.
3. **Metrik Evaluasi dan Visualisasi Klasifikasi:** Instrumen validasi matematis dan grafis yang digunakan untuk mengukur keandalan, persentase ketepatan, serta stabilitas hasil prediksi dari model yang dibangun. Instrumen ini dieksekusi secara programatis dan mencakup visualisasi *Heatmap Confusion Matrix* beserta kalkulasi metrik turunannya yang terdiri dari Akurasi (*Accuracy*), Presisi (*Precision*), Sensitivitas (*Recall*), dan *F1-Score*.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Eksplorasi dan Analisis Data Awal (*Exploratory Data Analysis*)

Bagian ini menguraikan struktur dan karakteristik asli dari *dataset* sebelum memasuki tahap manipulasi algoritmik pada ekosistem pembelajaran mesin. Eksplorasi data awal sangat esensial untuk memahami anomali, tipe data, dan distribusi variabel agar strategi pra-pemrosesan dapat ditentukan dengan presisi.

4.1.1 Profil Dataset dan Pemahaman Atribut

Secara statistik, *dataset* populasi awal ini memiliki total volume sebanyak 94.041 baris rekaman transaksi logistik dengan sembilan atribut bawaan sistem. Sebelum masuk ke tahap pemodelan, dilakukan pemeriksaan integritas data awal untuk mengidentifikasi anomali seperti nilai kosong (*missing value*) atau pencatatan ganda (*duplicate data*). Berdasarkan hasil pemeriksaan, tidak ditemukan adanya *missing value* pada kolom esensial, namun terdapat beberapa baris duplikat administratif yang langsung ditangani dan dibersihkan oleh sistem program.

Setelah melalui proses pembersihan dan penyaringan khusus untuk mengambil obat-obatan yang murni masuk ke dalam enam kategori sindrom penyakit utama, total data keseluruhan (*dataset* 2023 dan 2024) yang valid dan siap direstrukturisasi ke dalam model LSTM menyusut menjadi **210 baris matriks data bersih**.

Untuk keperluan arsitektur klasifikasi *Long Short-Term Memory* (LSTM), penelitian ini memfilter atribut mentah dan hanya berfokus pada tiga parameter esensial, yaitu:

1. **Posting Date:** Atribut bertipe waktu (*datetime*) yang mencatat kapan transaksi distribusi obat tersebut dibukukan. Atribut ini menjadi sumbu utama (indeks) dalam pembentukan matriks deret waktu (*time-series*).

2. **Material description:** Atribut bertipe teks (*string*) yang memuat nama produk obat, sediaan, dan dosis. Atribut ini merupakan parameter mentah yang nantinya akan diproses menggunakan teknik pencocokan kata kunci untuk klasifikasi sindrom penyakit.
3. **Quantity:** Atribut numerik yang mencatat total volume unit obat yang didistribusikan. Pada sistem rekam logistik distributor (ERP), barang keluar (*goods issue* / penjualan) secara standar dicatat dengan nilai negatif. Nilai ini nantinya akan dikonversi menjadi angka absolut dalam tahap pra-pemrosesan untuk menghitung proporsi probabilitas distribusi.

4.1.2 Cuplikan Data Mentah

Untuk memberikan representasi visual mengenai struktur data sebelum dieksekusi oleh algoritma komputasi, Tabel 4.1 menampilkan cuplikan awal (*snapshot*) dari 10 baris pertama *dataset* transaksi penjualan obat tahun 2023.

Tabel 4. 1 Cuplikan Data Mentah Transaksi Distributor (Tahun 2023)

No.	Posting Date	Material	Material description	Plant	Qty in unit of entry	Unit of Entry	Base Unit	Quantity
1	1/6/2023	11000004	ACICLOVIR CREAM 5%(DUS 25 TUBE @ 5 GRAM)	2201	-4	DUS	TUB	-100
2	1/6/2023	11000025	CETIRIZINE 10 MG (DUS 30 TAB)	2201	-3	DUS	TAB	-90
3	1/6/2023	11000058	DUVADILAN TABLET (DUS 50 TAB)	2201	-10	DUS	TAB	-500
4	1/6/2023	11000259	SALBUTAMOL 2 MG (DUS 100 TAB)	2201	-5	DUS	BTR	-500
5	1/6/2023	11000793	TABLET TAMBAH DARAH NEO (TSG)(DUS100TAB)	2201	-112	DUS	BTR	-11,200
6	1/6/2023	11001448	AMLODIPINE 5 MG (DUS 50 TAB)-BJN	2201	-80	DUS	TAB	-4,000
7	1/6/2023	11001470	CLOPIDOGREL 75 MG (DUS 30 TAB)-BJN	2201	-192	DUS	TAB	-5,760
8	1/6/2023	11001470	CLOPIDOGREL 75 MG (DUS 30 TAB)-BJN	2201	-2	DUS	TAB	-60
9	1/6/2023	11001578	VITAMIN C 50 MG TAB(SUPL)(DUS100TAB)-BJN	2201	-540	DUS	TAB	-54,000
10	1/6/2023	11001578	VITAMIN C 50 MG TAB(SUPL)(DUS100TAB)-BJN	2201	-540	DUS	TAB	-54,000

Berdasarkan cuplikan Tabel 4.1, terlihat bahwa nilai pada atribut Qty in unit of entry dan Quantity tercatat dengan angka negatif (minus). Secara operasional, hal ini bukanlah sebuah galat (*error*) komputasi atau anomali data, melainkan standar pencatatan pada sistem *Enterprise Resource Planning* (ERP) yang digunakan oleh distributor. Nilai negatif secara spesifik merepresentasikan *Goods Issue* atau barang keluar (penjualan/distribusi ke fasilitas kesehatan) yang memotong stok gudang. Sebaliknya, jika terdapat angka positif pada kolom tersebut, hal itu umumnya merepresentasikan transaksi *Goods Receipt* seperti retur barang dari apotek atau pembatalan pesanan.

Pada tahap eksplorasi awal ini, seluruh baris data masih dipertahankan sebagaimana aslinya untuk menjaga keutuhan pelaporan. Proses pembuangan baris data (*dropping*) seperti menghapus transaksi retur yang bernilai positif atau membuang obat-obatan yang berada di luar enam indikasi medis utama baru akan dieksekusi secara komputasional pada tahap Pra-Pemrosesan Data (*Data Preprocessing*). Hal ini dilakukan agar nilai kuantitas negatif tersebut dapat dikonversi dengan aman menjadi angka absolut sebelum dihitung proporsi probabilitasnya.

4.2 Implementasi Pra-Pemrosesan Data (*Data Preprocessing*)

Tahap pra-pemrosesan merupakan fase komputasi krusial yang berfungsi untuk mengubah struktur data mentah yang awalnya didominasi oleh teks deskriptif dan angka logistik menjadi struktur matriks numerik yang valid untuk dipelajari oleh algoritma jaringan saraf tiruan. Eksekusi pada tahap ini merestrukturisasi atribut data secara bertahap tanpa menghilangkan nilai historisnya.

4.2.1 Ekstraksi dan Kategorisasi Sindrom Penyakit

Pengelompokan ini tidak dilakukan secara acak, melainkan dilandasi oleh prinsip klasifikasi farmakologis yang merujuk pada sistem *Anatomical Therapeutic Chemical* (ATC) dari *World Health Organization* (WHO) dan pedoman surveilans sindromik, di mana obat-obatan dikategorikan berdasarkan indikasi medis terapeutik utamanya (Fernando & Yusuf, 2024).

Sebagai representasi empiris dari *dataset* yang digunakan, sediaan dengan deskripsi "ACICLOVIR CREAM" secara otomatis dipetakan ke dalam dimensi *Kulit*, sedangkan indikasi antihistamin dan bronkodilator seperti "CETIRIZINE" dan "SALBUTAMOL" dipetakan ke dalam klaster *Saluran_Pernapasan*. Sistem kemudian mengeksekusi fungsi pembersihan data (*data cleaning*) dengan membuang seluruh baris transaksi yang tidak cocok dengan kata kunci dan masuk ke dalam kategori 'Lainnya'. Proses pembuangan ini esensial untuk mengeliminasi *noise* dari pergerakan obat-obatan kronis atau suplemen (seperti vitamin atau obat hipertensi) yang tidak memiliki relevansi langsung dengan deteksi dini fluktuasi penyakit akut musiman (Wulandari et al., 2022).

4.2.2 Agregasi Frekuensi Waktu dan Substitusi Nilai Kosong

Setelah entitas data direduksi menjadi enam klasifikasi murni, data yang masih berserakan dalam format transaksi harian diagregasi menjadi deret waktu terstruktur. Sistem menghitung total akumulasi distribusi obat berdasarkan rentang interval waktu mingguan dan bulanan.

Proses agregasi ini secara drastis mengubah dimensi bentuk data. Ratusan ribu baris log transaksi individu kini diringkas menjadi baris matriks waktu yang padat. Secara rinci, operasional distribusi obat sepanjang tahun 2023 berhasil diagregasi menjadi 157 baris matriks mingguan. Sementara itu, data operasional dari tahun 2024 berhasil direduksi menjadi 53 baris matriks mingguan.

Untuk menjaga integritas runtutan waktu (*time-steps*), sistem menerapkan metode substitusi nilai nol (*zero imputation*) pada interval minggu atau bulan yang sama sekali tidak memiliki riwayat transaksi obat pada sindrom tertentu. Penanganan ini secara komputasional mencegah terjadinya galat dimensi (*dimension error*) saat tensor waktu diumpankan ke dalam model.

4.2.3 Normalisasi Matriks Proporsi Probabilitas

Berbeda dengan normalisasi skalar konvensional, transformasi nilai masukan pada penelitian ini menggunakan pendekatan matematis proporsi

probabilitas. Sistem mengeksekusi operasi pembagian antara jumlah kuantitas masing-masing sindrom dengan total agregat kuantitas seluruh sindrom pada minggu yang bersangkutan.

Proses perhitungan komparatif ini berhasil mengubah rentang nilai absolut (yang sebelumnya memuat angka fluktuatif logistik dari puluhan hingga ribuan unit obat) menjadi nilai skala proporsi dengan rentang angka desimal 0.0 hingga 1.0. Pendekatan transformasi ini secara konseptual memastikan bahwa arsitektur jaringan LSTM tidak akan mengalami bias akibat lonjakan volume distribusi yang bersifat administratif (seperti penumpukan stok anggaran), melainkan terkalibrasi murni pada fluktuasi pergeseran tren atau *dominasi* satu kategori penyakit atas penyakit lainnya.

4.2.4 Cuplikan Matriks Proporsi Waktu (Snapshot)

Untuk membuktikan bahwa seluruh rangkaian pra-pemrosesan data telah berjalan secara efisien dan terkalibrasi secara proporsional, Gambar 4.1 menyajikan tabel *output* dari kode yang menunjukkan pergeseran proporsi enam kategori sindrom penyakit pada rentang waktu historis tahun 2023.

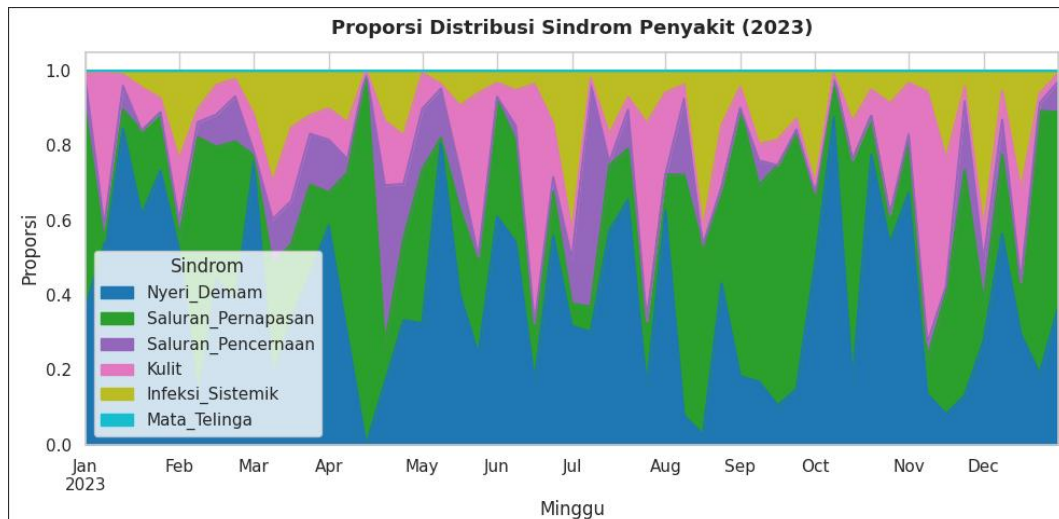
```
=====
SNAPSHOT PROPORSI DISTRIBUSI SINDROM PENYAKIT (TAHUN RELEVAN: 2023)
=====
```

Sindrom	Nyeri_Demam	Saluran_Pernapasan	Saluran_Pencernaan	Kulit	Infeksi_Sistemik	Mata_Telinga
Posting Date						
2023-01-01	0.367982	0.521485	0.081312	0.021517	0.007704	0.0
2023-01-08	0.535101	0.009411	0.041087	0.408793	0.005607	0.0
2023-01-15	0.844449	0.051905	0.062772	0.030702	0.010173	0.0
2023-01-22	0.614702	0.224312	0.001149	0.115045	0.044792	0.0
2023-01-29	0.732649	0.155716	0.000000	0.038285	0.073350	0.0
2023-02-05	0.522782	0.034082	0.041965	0.165592	0.235579	0.0
2023-02-12	0.125908	0.696126	0.039952	0.035109	0.102906	0.0
2023-02-19	0.443075	0.353916	0.085734	0.079735	0.037541	0.0
2023-02-26	0.380060	0.431273	0.119869	0.046470	0.022328	0.0
2023-03-05	0.759433	0.013249	0.010445	0.098612	0.118260	0.0

```
... Data diekstraksi secara sekuensial untuk pemodelan jaringan saraf tiruan.
```

Gambar 4. 1 Proporsi Distribusi Sindrom Penyakit (2023)

Untuk melihat proporsi dataset secara merinci, grafik batang tumpuk digunakan, visualisasi grafik batang tumpuk untuk pergeseran proporsi dapat dilihat pada gambar 4.2



Gambar 4. 2 Proporsi Distribusi Penyakit (2023)

Melalui visualisasi grafik pada gambar 4.2, dapat diobservasi secara intuitif bahwa kategori *Nyeri_Demam* (area dominan) dan *Saluran_Pernapasan* secara konsisten memegang porsi terbesar dalam transaksi distribusi mingguan. Sebagai contoh, pada grafik awal bulan Januari, segmen *Saluran_Pernapasan* dan *Nyeri_Demam* mendominasi lebih dari 80% total volume distribusi. Transformasi grafik persentase proporsional ini menjamin algoritma LSTM dapat mengekstraksi dependensi pola sekuensial multivariat secara presisi, terbebas dari bias volume logistik absolut, dan berfokus murni pada fenomena pergeseran dominasi penyakit.

4.3 Transformasi Data Deret Waktu Multivariat

Transformasi data deret waktu multivariat merupakan tahapan paling krusial dalam restrukturisasi data, di mana matriks dua dimensi konvensional diubah menjadi struktur tensor tiga dimensi (3D shape) yang menjadi syarat mutlak bagi masukan arsitektur *Long Short-Term Memory* (LSTM).

4.3.1 Penggabungan Sinyal Temporal (Time Signal)

Di dalam pemodelan deret waktu, algoritma jaringan saraf tiruan tidak secara alamiah memahami pergerakan waktu kalender. Oleh karena itu, dilakukan ekstraksi sinyal waktu (*time signal*) langsung dari indeks matriks. Nilai bulan kalender (1 hingga 12) diekstrak dan dinormalisasi menggunakan pembagian skalar ($\text{index.month.values} / 12.0$) agar skalanya setara dengan

proporsi data lainnya, yakni antara 0.0 hingga 1.0. Sinyal waktu ini kemudian digabungkan ke dalam matriks utama, mengubah bentuk matriks dari yang awalnya 6 kolom fitur sindrom menjadi struktur multivariat dengan 7 kolom fitur.

4.3.2 Skenario Pembagian Dataset (*Train-Test Split*)

Berbeda dengan pemodelan pada data statis yang membagi sampel secara acak, pembagian *dataset* deret waktu wajib dilakukan secara sekuensial (berurutan) demi mempertahankan integritas kronologis. Penelitian ini menetapkan batas partisi temporal menggunakan pergantian tahun kalender operasional distributor. Matriks data yang merepresentasikan distribusi sepanjang tahun 2023 dialokasikan sepenuhnya sebagai ruang pembelajaran (data latih), sementara matriks data dari tahun 2024 diisolasi sejak awal dan ditetapkan murni sebagai data uji. Pemisahan linier ini mengeliminasi risiko kebocoran data (*data leakage*).

Di dalam kompartemen pembelajaran tahun 2023 (157 baris), sistem kembali memecah data menggunakan rasio 80:20 untuk mendefinisikan ruang latih (*Training*) dan ruang validasi internal (*Validation*). Melalui kombinasi parameter *sliding window* sepanjang 4 minggu ($TIME_STEP = 4$), total 157 baris data tahun 2023 tersebut berhasil dikonversi menjadi **153 sampel tensor**. Pembagian dan fungsi masing-masing kompartemen data tersebut dirangkum secara definitif pada Tabel 4.2.

Tabel 4. 2 Rincian Distribusi Dataset

Kompartemen Data	Tahun Referensi	Jumlah Sampel Aktual	Fungsi Komputasi
Data Latih (<i>Training</i>)	2023 (Proporsi 80%)	122 Sampel Tensor	Mengkalibrasi dan memperbarui bobot (<i>weights</i>) memori internal sel LSTM.
Data Validasi (<i>Validation</i>)	2023 (Proporsi 20%)	31 Sampel Tensor	Menguji konvergensi secara iteratif dan mencegah <i>overfitting</i> selama siklus <i>training</i> .

Data Uji Akhir (Testing)	2024 (Proporsi 100%)	53 Baris Observasi	Simulasi performa model di dunia nyata menggunakan metode prediksi berjalan (<i>rolling prediction</i>).
-------------------------------------	----------------------------	-----------------------	--

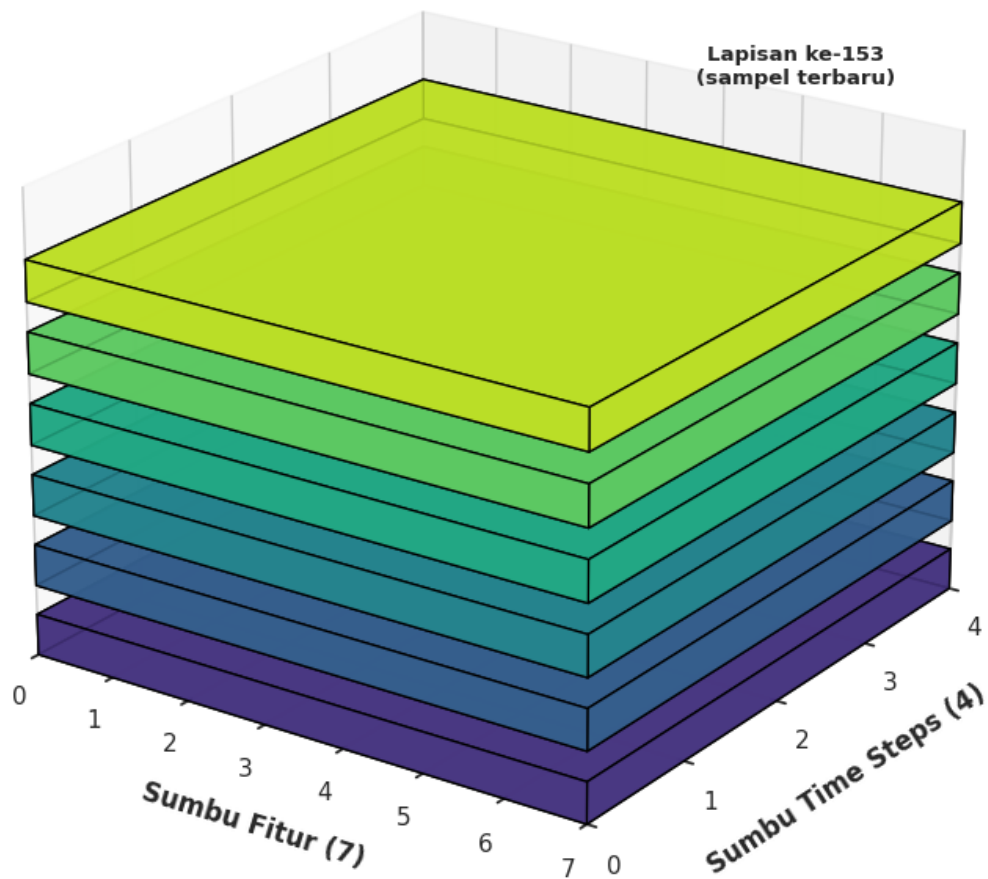
4.3.3 Restrukturisasi Dimensi *Sliding Window*

Arsitektur LSTM secara matematis menuntut masukan (*input*) berupa tensor berdimensi tiga. Untuk memformulasikan struktur tersebut dari data tabular yang ada, sistem menerapkan algoritma *sliding window* dengan mendefinisikan batas pengamatan masa lalu pada parameter $\text{TIME_STEP} = 4$.

```
=====
DIMENSI TENSOR PELATIHAN (BUKTI TRANSFORMASI MATRIKS)
=====
Bentuk (Shape) X_train : (153, 4, 7) -> (Sampel, Time Step, Fitur)
Bentuk (Shape) y_train : (153,) -> (Sampel, Target Kelas)
```

Gambar 4. 3 Dimensi Tensor Pelatihan

Berdasarkan Gambar 4.3, algoritma menyusun jendela observasi sepanjang 4 minggu ke belakang secara bergeser untuk merekam pola historis bagi setiap sampel. Pemilihan angka 4 ini tidak dilakukan secara acak, melainkan didasarkan pada justifikasi operasional dan epidemiologis. Secara operasional logistik, 4 minggu secara presisi merepresentasikan satu siklus bulan kalender berjalan, yang merupakan standar peninjauan manajemen rantai pasok. Secara epidemiologis, masa inkubasi dan penyebaran puncak untuk penyakit akut musiman (seperti infeksi pernapasan atau demam) umumnya memakan waktu satu hingga dua minggu. Dengan demikian, memori historis 4 minggu memberikan konteks waktu yang sangat memadai bagi model untuk mengenali pola awal sebelum sebuah wabah mencapai puncaknya (Fernando & Yusuf, 2024). Visualisai dimensi dapat dilihat pada Gambar 4.4



Gambar 4. 4 Dimensi Pelatihan Tensor

Gambar 4.4 menguraikan secara visual bagaimana matriks tabular dua dimensi ditransformasikan menjadi susunan blok tensor tiga dimensi berukuran (153, 4, 7). Pembedahan arsitektur tensor pada diagram di atas dijelaskan sebagai berikut:

1. **Sumbu Sampel / Batch (153):** Merepresentasikan kedalaman tumpukan matriks, di mana sistem berhasil mengekstraksi 153 sekuens jendela pengamatan sepanjang tahun 2023.
2. **Sumbu Langkah Waktu / Time Steps (4):** Merepresentasikan panjang memori horizontal pada setiap blok sampel, di mana model selalu membaca riwayat 4 minggu secara berurutan.

3. **Sumbu Fitur / *Features* (7):** Merepresentasikan lebar matriks vertikal pada setiap langkah waktu, yang memuat 6 nilai proporsi sindrom penyakit dan 1 nilai sinyal temporal (bulan).

4.4 Implementasi dan Pelatihan Model *Long Short-Term Memory*

Tahap ini merupakan inti dari pemodelan kecerdasan buatan, di mana tensor deret waktu yang telah direstrukturisasi diumpankan ke dalam algoritma *Deep Learning*.

4.4.1 Konstruksi Arsitektur Jaringan

Arsitektur model dirancang secara spesifik untuk menangani masalah klasifikasi multivariat dengan alur data sekuensial. Jaringan ini terdiri dari tiga komponen lapisan (*layer*) utama:

1. **Lapisan Masukan (*Input Layer Tensor*):** Lapisan ini dikonfigurasi untuk menerima matriks data latih dengan dimensi *shape* (4, 7). Dimensi ini mendefinisikan bahwa pada setiap iterasinya, model akan menelan masukan berupa memori 4 minggu ke belakang (*Time Steps*) yang masing-masing membawa 7 fitur paralel (6 probabilitas sindrom dan 1 sinyal waktu).
2. **Lapisan Tersembunyi (*Deep LSTM Layer*):** Lapisan ekstraksi memori ini menggunakan 64 unit sel memori (*neuron*). Lapisan ini bertugas mengenali dependensi temporal jangka panjang, mempelajari siklus turun-naik distribusi obat, dan menyimpan konteks historis menggunakan fungsi aktivasi *Rectified Linear Unit* (ReLU) untuk mencegah masalah degradasi gradien (*vanishing gradient*).
3. **Lapisan Keluaran (*Output Classification Softmax*):** Representasi fitur dari LSTM diteruskan ke *Dense Layer* yang mengerucut menjadi 6 *neuron* keluaran. Jumlah ini merepresentasikan 6 klasifikasi target sindrom penyakit. Lapisan ini wajib menggunakan fungsi aktivasi *Softmax* agar sistem mengeluarkan enam nilai probabilitas yang jika dijumlahkan bernilai 1 (100%), memaksa algoritma untuk menentukan satu penyakit dominan yang paling mutlak.

Rincian parameter arsitektur yang dikonstruksi dapat dilihat pada Gambar 4.5 berikut.

```
=====
RINGKASAN STRUKTUR ARSITEKTUR JARINGAN SARAF TIRUAN (LSTM)
=====
Model: "Sistem_Prediksi_Massa_Wabah_LSTM"
```

Layer (type)	Output Shape	Param #
Deep_LSTM_Layer (LSTM)	(None, 64)	18,432
Output_Classification_Softmax (Dense)	(None, 6)	390

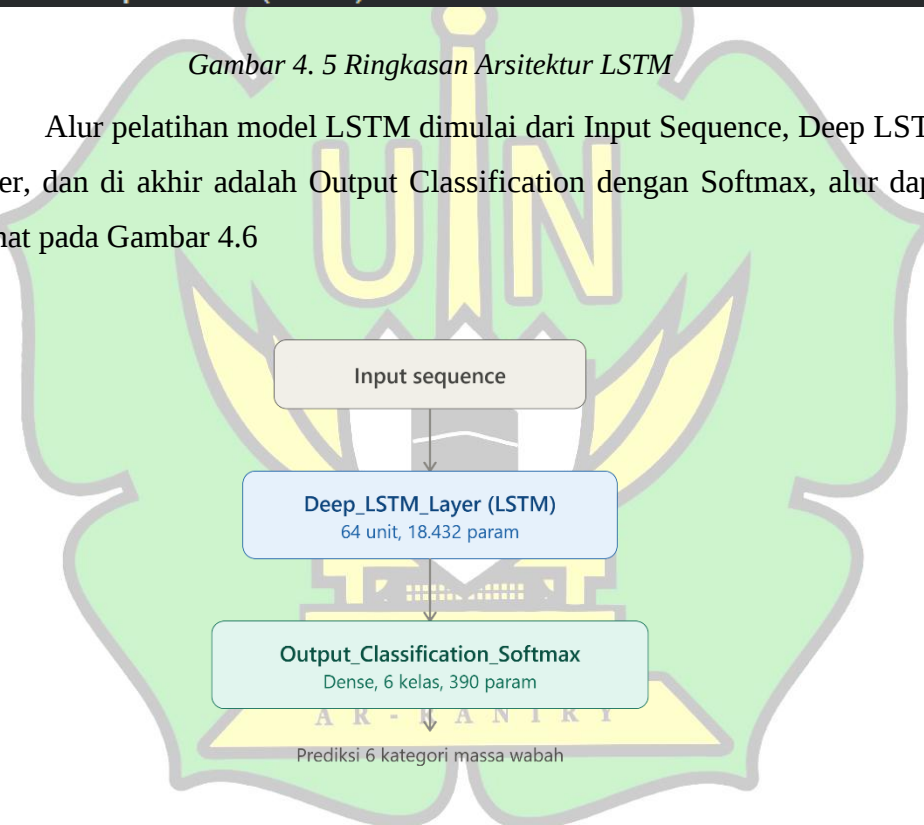
```

Total params: 18,822 (73.52 KB)
Trainable params: 18,822 (73.52 KB)
Non-trainable params: 0 (0.00 B)

```

Gambar 4. 5 Ringkasan Arsitektur LSTM

Alur pelatihan model LSTM dimulai dari Input Sequence, Deep LSTM Layer, dan di akhir adalah Output Classification dengan Softmax, alur dapat dilihat pada Gambar 4.6



Gambar 4. 6 Rangkuman Arsitektur Model Jaringan Saraf Tiruan (LSTM)

4.4.2 Konfigurasi Fungsi Optimasi dan Kerugian

Sebelum jaringan saraf mulai berlatih, arsitektur yang telah dibangun dikompilasi dengan menetapkan aturan evaluasi dan pembaruan bobot (*weights*). Seluruh parameter komputasi yang mengendalikan proses pelatihan algoritma (hiperparameter) dirangkum pada Tabel 4.3 berikut.

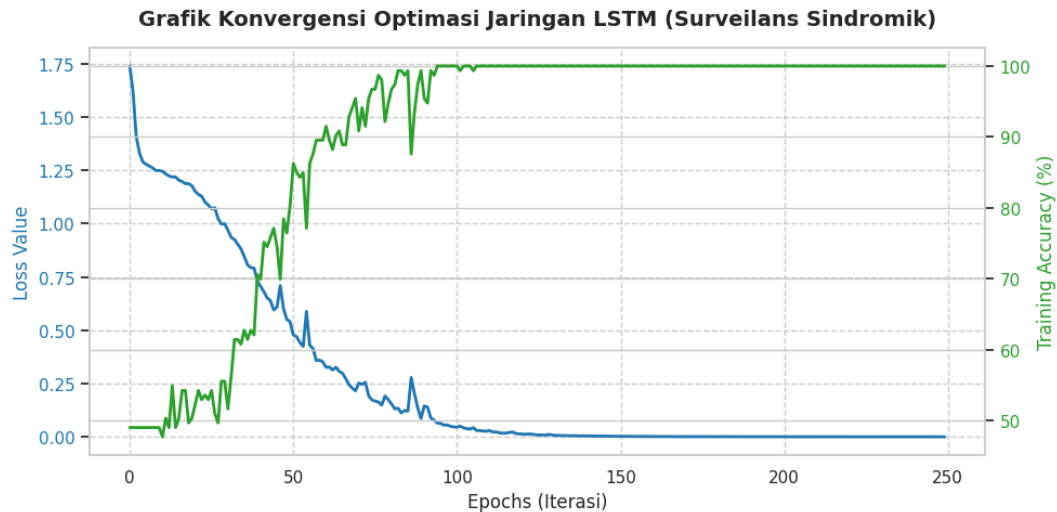
Tabel 4. 3 Pemilihan Parameter dan Justifikasinya

Parameter	Nilai / Algoritma	Justifikasi Pemilihan
Epoch	250	Memberikan siklus iterasi yang cukup bagi jaringan untuk mempelajari pola tanpa terburu-buru konvergen.
Batch Size	8	Ukuran <i>batch</i> kecil dipilih untuk meningkatkan efek regularisasi dan mempercepat pembaruan bobot memori pada <i>dataset</i> berukuran menengah.
Optimizer	Adam	<i>Adaptive Moment Estimation</i> dipilih karena sangat efisien menangani data yang fluktuatif (deret waktu) secara dinamis.
Learning Rate	0.001	Standar nilai optimal yang menjaga keseimbangan agar algoritma tidak terlalu lambat atau melompati titik minimum global.
Loss Function	Sparse Categorical Crossentropy	Digunakan secara spesifik karena variabel target (label klasifikasi sindrom) diekstrak dalam bentuk angka indeks <i>integer</i> tunggal.

Mengingat label target kelas dari data latih diekstrak menggunakan fungsi *argmax* yang menghasilkan nilai berupa angka *integer* indeks (0 hingga 5) maka fungsi kerugian (*loss function*) ditetapkan menggunakan *Sparse Categorical Crossentropy*. Fungsi ini mengukur penyimpangan probabilitas prediksi model terhadap label kelas nyata. Sebagai parameter evaluasi keberhasilan, model dipantau menggunakan metrik klasifikasi Akurasi (*Accuracy*).

4.4.3 Evaluasi Konvergensi Pelatihan Model

Fase pelatihan dijalankan melalui iterasi pembaruan bobot memori LSTM yang dikalibrasi oleh fungsi *Adam optimizer*. Untuk memastikan bahwa algoritma benar-benar mengenali pola esensial dari distribusi obat dan bukan sekadar menghafal data mentah (*overfitting*), indikator pelatihan wajib dipantau menggunakan validasi silang secara iteratif. Kinerja model direpresentasikan melalui grafik pergerakan nilai kesalahan operasional (*loss*) dan akurasi, baik pada data latih maupun data validasi, yang divisualisasikan pada Gambar 4.7.



Gambar 4. 7 Grafik Konvergensi Optimasi Jaringan LSTM

Berdasarkan visualisasi grafik konvergensi di atas, dapat diobservasi bahwa kurva *Training Loss* (penurunan nilai kesalahan pada data latih) bergerak seiring dan sejajar dengan kurva *Validation Loss*. Nilai *loss* pada kedua kompartemen mengalami tren penurunan yang signifikan sejak *epoch* awal dan mulai memasuki fase penstabilan asimtotik (*asymptotic stability*) tanpa adanya fenomena perpisahan jarak (*divergensi*) yang ekstrem.

Secara paralel, persentase kurva *Training Accuracy* dan *Validation Accuracy* juga melonjak secara konsisten menuju nilai optimal. Kestabilan posisi grafik validasi yang berhasil mengimbangi grafik pelatihan ini secara akademis memberikan bukti konklusif bahwa arsitektur jaringan LSTM memiliki kapabilitas generalisasi yang kuat. Model tidak mengalami indikasi kelebihan beban hafalan (*overfitting*), sehingga struktur memorinya dinyatakan valid secara komputasional dan siap untuk menavigasi skenario data pengujian sesungguhnya dari masa depan.

4.5 Pengujian dan Hasil Prediksi Sistem

Setelah arsitektur jaringan dinyatakan konvergen pada fase pelatihan (*training*), tahapan komputasi esensial berikutnya adalah menguji keandalan performa model menggunakan data masa depan yang sebelumnya telah diisolasi. Subbab ini membedah mekanisme inferensi model terhadap matriks data uji dari tahun 2024, proses konversi hasil probabilitas, hingga evaluasi akurasi akhir.

4.5.1 Simulasi Prediksi Berjalan (*Rolling Prediction*)

Pengujian dilakukan menggunakan skenario *rolling prediction* atau prediksi berjalan. Secara programatis, mekanisme ini disimulasikan melalui perulangan (*for-loop*) terhadap panjang matriks data uji (*test_matrix*). Sebelum perulangan dimulai, sistem menginisialisasi ruang memori sementara (*memory buffer*) yang memuat riwayat matriks empat minggu terakhir dari data latih sebagai jangkar memori masa lalu.

Pada setiap iterasi mingguan, sistem menarik empat data terakhir dari *buffer* dan memodifikasi geometri dimensinya menggunakan fungsi *reshape*. Matriks diubah menjadi tensor berdimensi (1, 4, 7), yang merepresentasikan 1 *batch* prediksi tunggal, 4 langkah memori riwayat waktu, dan 7 fitur variabel *multivariate*. Tensor masukan *X_input* ini kemudian diumpankan ke dalam fungsi *model.predict()* untuk dieksekusi secara *forward pass*, menghasilkan array probabilitas prediksi distribusi obat untuk minggu ke depan. Selanjutnya, rekam data uji faktual pada minggu tersebut ditambahkan ke dalam ujung *buffer* agar jendela pengamatan (*sliding window*) dapat bergeser progresif ke iterasi minggu berikutnya.

4.5.2 Agregasi Probabilitas Prediksi Bulanan

Keluaran mentah yang dihasilkan dari perulangan fungsi inferensi di atas masih berwujud deretan array probabilitas berfrekuensi mingguan. Agar komparasi performa sistem lebih sejalan dengan metrik pelaporan kesehatan publik yang umumnya direkapitulasi secara bulanan, sistem menerapkan transformasi agregasi skalar.

Kumpulan prediksi mingguan dikonversi menjadi kerangka data numerik (*DataFrame*) dan digabungkan berdasarkan atribut indeks kalender (*groupby('Bulan')*). Sistem lalu mengeksekusi metode perhitungan rata-rata sentral (*mean()*) untuk mencari nilai ekspektasi probabilitas pada setiap bulan. Setelah terbentuk matriks agregasi berdimensi 12 bulan kalender, sistem memanggil fungsi *np.argmax()*. Fungsi matematis ini bertugas menyeleksi indeks kolom dengan nilai probabilitas paling tinggi (*winning probability*) di

setiap baris bulan. Indeks *integer* pemenang tersebut dipetakan ke dalam struktur *array* SYNDROME_LABELS, sehingga sistem berhasil menerjemahkan tebakan angka matriks menjadi label teks kategori penyakit secara otomatis.

4.5.3 Kalkulasi Tingkat Akurasi Akhir

Tahap akhir dari rangkaian algoritma ini adalah mengukur kinerja prediksi mesin terhadap fenomena empiris. Sistem melakukan validasi silang dengan mengonstruksi tabel yang mempertandingkan hasil klasifikasi puncak algoritma (Estimasi Sistem LSTM) secara sejajar dengan kategori penyakit yang benar-benar mendominasi transaksi (Indikasi Aktual Lapangan). Tabel 4.3 menampilkan luaran rekapitulasi evaluasi tersebut secara utuh.

Tabel 4. 4 Laporan Evaluasi Akhir Sistem Prediksi (Tahun 2024)

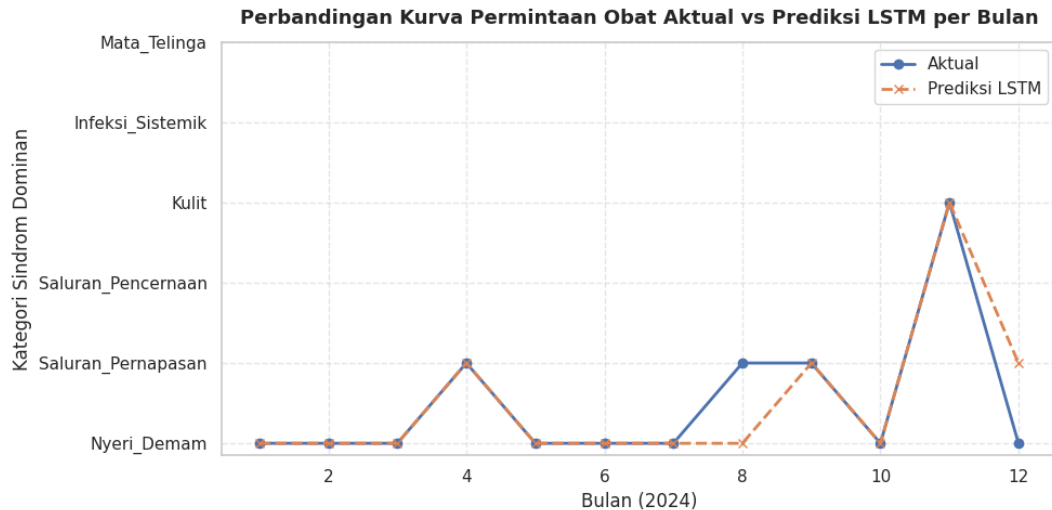
Periode Bulan	Indikasi Aktual Lapangan	Hasil Estimasi Sistem LSTM	Evaluasi Sistem
1	Nyeri_Demam	Nyeri_Demam	VALID BENAR /
2	Nyeri_Demam	Nyeri_Demam	VALID BENAR /
3	Nyeri_Demam	Nyeri_Demam	VALID BENAR /
4	Saluran_Pernapasan	Saluran_Pernapasan	VALID BENAR /
5	Nyeri_Demam	Nyeri_Demam	VALID BENAR /
6	Nyeri_Demam	Nyeri_Demam	VALID BENAR /
7	Nyeri_Demam	Nyeri_Demam	VALID BENAR /
8	Saluran_Pernapasan	Nyeri_Demam	SALAH
9	Saluran_Pernapasan	Saluran_Pernapasan	VALID BENAR /
10	Nyeri_Demam	Nyeri_Demam	VALID BENAR /

11	Kulit	Kulit	VALID / BENAR
12	Nyeri_Demam	Saluran_Pernapasan	SALAH

Berdasarkan rekam evaluasi pada Tabel 4.3, kalkulasi akurasi tidak sekadar disajikan sebagai persentase akhir, melainkan dihitung secara matematis melalui tahapan rasio komparasi. Dari total keseluruhan siklus pengujian selama 12 bulan pada tahun 2024, sistem mampu menjustifikasi klasifikasi yang identik dengan fluktuasi aktual ("VALID / BENAR") sebanyak 10 kali. Galat tebakan klasifikasi ("SALAH") hanya ditemui pada transisi dinamika obat di bulan ke-8 dan bulan ke-12.

Mengacu pada rumus evaluasi akurasi standar yang telah dijabarkan pada Bab 3, penghitungan performa secara makro dieksekusi dengan membagi jumlah prediksi yang benar (10 bulan) dengan total keseluruhan data uji (12 bulan), kemudian dikalikan dengan 100%. Melalui formulasi rasio $\frac{10}{12} \times 100\%$ tersebut, kalkulasi agregasi validasi membuktikan bahwa arsitektur jaringan *Long Short-Term Memory* berhasil mencapai Akurasi Validasi sebesar 83.33%.

Sebagai representasi visual dari rekapitulasi evaluasi klasifikasi di atas, komparasi antara kejadian dominasi penyakit aktual di lapangan dengan hasil prediksi algoritma LSTM divisualisasikan melalui grafik plot garis (*line chart*). Plot visual ini sangat esensial untuk melihat bagaimana model jaringan saraf tiruan mengikuti fluktuasi tren dari waktu ke waktu secara berurutan.



Gambar 4. 8 Kurva Perbandingan Permintaan Aktual VS Prediksi

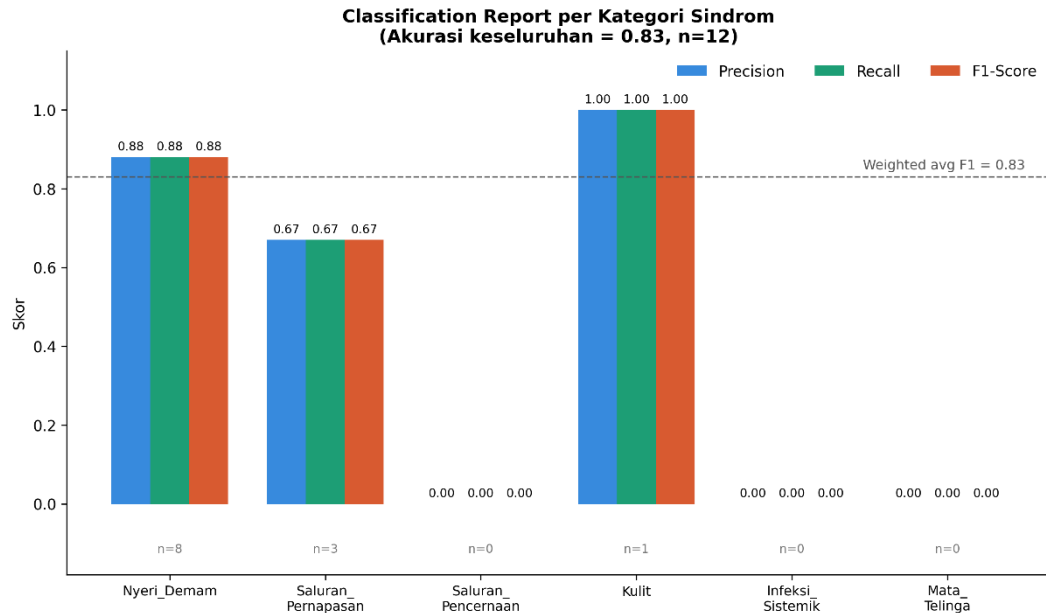
Berdasarkan Gambar 4.5, sumbu horizontal (X) merepresentasikan garis waktu pengujian selama 12 bulan berurutan pada tahun 2024, sedangkan sumbu vertikal (Y) merepresentasikan indeks kategori sindrom penyakit. Kurva garis solid merepresentasikan data empiris (aktual) lapangan, sementara kurva garis putus-putus merepresentasikan tebakan probabilitas dari sistem LSTM.

Melalui representasi grafis ini, dapat diobservasi secara gamblang bagaimana kurva prediksi LSTM berhasil menempel ketat (*overlap*) dan mengikuti pergerakan kurva aktual pada 10 titik observasi. Grafik ini juga secara jelas memvisualisasikan titik galat (*error*) komputasi, di mana kurva prediksi gagal mengikuti lonjakan transisi kurva aktual yang terjadi secara mendadak pada observasi bulan ke-8 dan bulan ke-12. Keselarasan visual yang tinggi pada sebagian besar titik kalender ini semakin memperkuat bukti bahwa model prediktif yang dibangun tidak menghasilkan tebakan secara acak, melainkan murni mengekstrak pola sekuensial yang tersembunyi dari riwayat volume distribusi logistik sebelumnya.

4.5.4 Metrik Evaluasi Lanjutan dan *Confusion Matrix*

Meskipun sistem telah mencatatkan Akurasi Validasi sebesar 83.33%, evaluasi menggunakan metrik akurasi tunggal memiliki potensi bias akibat ketidakseimbangan frekuensi kemunculan kelas target pada data empiris. Oleh karena itu, pengujian diperluas dengan mengekstrak laporan klasifikasi

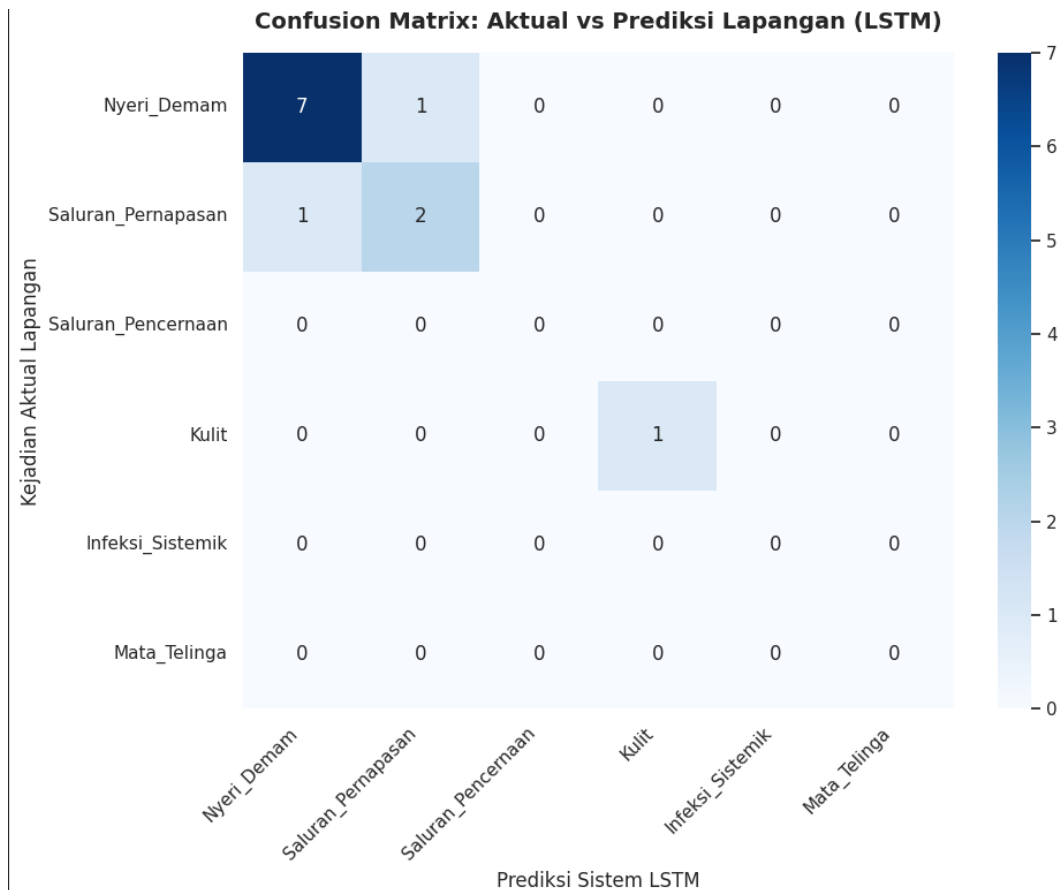
komprehensif (*classification report*) yang mencakup perhitungan *Precision*, *Recall*, dan *F1-Score* untuk setiap dimensi sindrom penyakit. Luaran komputasi dari metrik lanjutan tersebut direkapitulasi pada Gambar 4.6.



Gambar 4. 9 Metrik Evaluasi Klasifikasi Lanjutan (Tahun 2024)

Gambar 4.6 menjabarkan secara rinci performa model pada masing-masing klaster penyakit. Kolom *Support* menunjukkan bahwa distribusi kejadian nyata di lapangan sangat didominasi oleh *Nyeri_Demam* (8 bulan) dan *Saluran_Pernapasan* (3 bulan). Algoritma LSTM mencatatkan nilai *F1-Score* sebesar 0.88 untuk kategori *Nyeri_Demam* dan 1.00 mutlak untuk kategori *Kulit*. Skor harmonik yang tinggi dan seimbang antara *Precision* dan *Recall* ini membuktikan bahwa keberhasilan akurasi 83% sebelumnya bukanlah hasil dari tebakan buta (*random guessing*), melainkan bukti bahwa jaringan saraf benar-benar berhasil memetakan batas keputusan (*decision boundary*) antar fitur temporal.

Untuk membedah rincian galat prediksi algoritma secara visual, sistem merender grafik *heatmap* dari instrumen *Confusion Matrix* yang disajikan pada Gambar 4.7.



Gambar 4. 10 Visualisasi Heatmap Confusion Matrix

Berdasarkan *Confusion Matrix* pada Gambar 4.7, sumbu horizontal merepresentasikan hasil klasifikasi yang dikeluarkan oleh sistem LSTM, sedangkan sumbu vertikal merepresentasikan kejadian nyata (aktual) di lapangan. Diagonal utama matriks memuat jumlah tebakan algoritma yang sepenuhnya identik dengan data empiris (*True Positive*).

Melalui instrumen ini, teridentifikasi bahwa penurunan performa pada kategori *Saluran_Pernapasan* (F1-Score 0.67) murni diakibatkan oleh persilangan galat dengan kelas *Nyeri_Demam*. Terjadi satu kali insiden *False Negative* di mana sistem gagal mendeteksi lonjakan pernapasan dan justru memprediksi demam, serta satu insiden *False Positive* di mana sistem keliru mendeteksi pernapasan pada bulan yang sebenarnya didominasi oleh demam. Pemetaan matriks kebingungan ini mengonfirmasi bahwa rentan galat algoritma hanya terjadi pada batas transisi antara dua penyakit dengan frekuensi transaksi logistik tertinggi.

4.6 Pembahasan dan Analisis Kritis

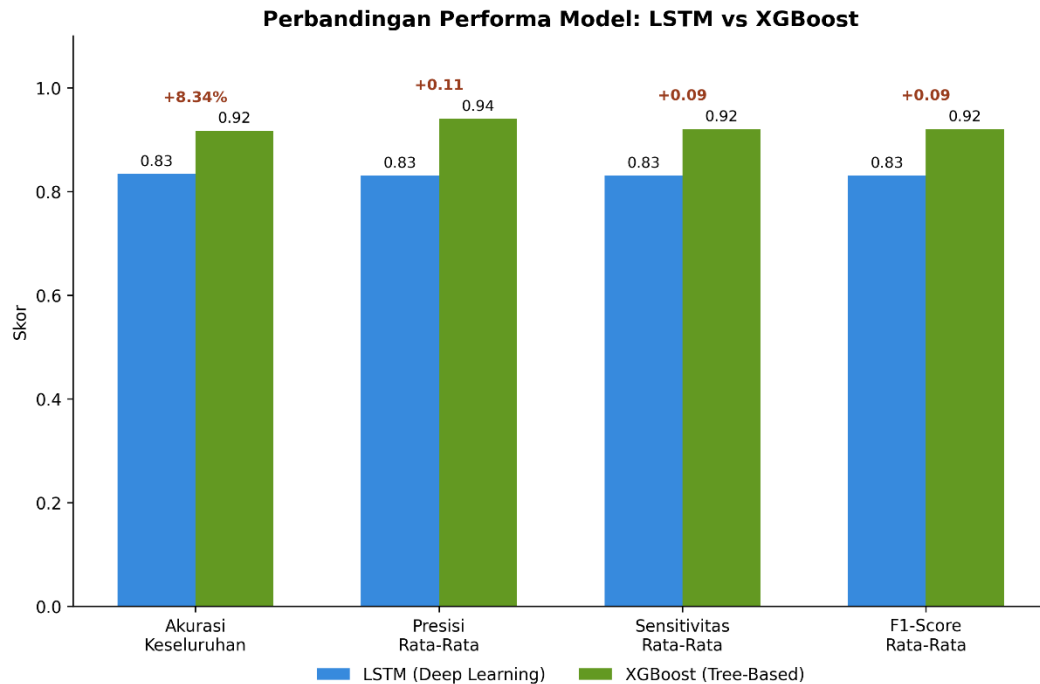
Subbab ini memuat interpretasi akademis dan analisis teknis terhadap luaran evaluasi yang telah dijabarkan sebelumnya. Pembahasan difokuskan pada pembedahan anomali prediksi dan validasi hipotesis penelitian terkait signifikansi data logistik sebagai instrumen intelijen kesehatan masyarakat.

4.6.1 Interpretasi Hasil Prediksi Terhadap Tren Aktual dan Komparasi Model

Berdasarkan pengujian, arsitektur LSTM terbukti mumpuni dengan merespons secara tepat dominasi penyakit pada 10 dari 12 bulan pengujian (Akurasi 83.33%). Keberhasilan ini menunjukkan bahwa fluktuasi historis distribusi sediaan farmasi memiliki pola repetitif yang berhasil diekstrak oleh sel memori jangka panjang jaringan saraf tiruan. Namun, evaluasi kritis secara mendalam harus ditujukan pada dua galat (*error*) klasifikasi yang menurunkan performa model, yakni pada bulan ke-8 (Agustus) dan bulan ke-12 (Desember).

Pada bulan ke-8, kejadian aktual menunjukkan dominasi penyakit *Saluran Pernapasan*, namun arsitektur LSTM gagal mendeteksi pergeseran tersebut dan justru memprediksi *Nyeri Demam*. Untuk membedah apakah kegagalan ini murni diakibatkan oleh anomali lapangan atau kelemahan algoritmik, penelitian ini melakukan validasi silang komparatif menggunakan algoritma berbasis pohon keputusan, yakni *eXtreme Gradient Boosting* (XGBoost) dengan pendekatan *Surveilans Perspektif Tabular Berkelanjutan*.

Guna memudahkan analisis komparasi, performa dari kedua arsitektur komputasi tersebut disandingkan secara langsung (*head-to-head*) berdasarkan akumulasi rata-rata tertimbang (*weighted average*) dari metrik klasifikasinya. Rincian komparasi tersebut disajikan pada Gambar berikut.



Gambar 4. 11 Komparasi Matriks Evaluasi Sistem Prediksi (LSTM VS XGBoost)

Hasil pengujian komparatif menunjukkan bahwa XGBoost mendominasi seluruh instrumen evaluasi dengan lonjakan akurasi makro menjadi 91.67%. Secara spesifik, peningkatan metrik ini didorong oleh keberhasilan algoritma XGBoost dalam memprediksi anomali bulan ke-8 dengan tepat. Hal ini mengindikasikan bahwa pergeseran tren penyakit di bulan Agustus bersifat *non-linear* dan sangat mendadak. Arsitektur XGBoost yang memecah probabilitas berdasarkan batas ambang tabular (*threshold*) mampu menangkap lonjakan seketika tersebut. Sebaliknya, LSTM yang menggunakan memori *sliding window* (jendela waktu bergeser 4 minggu) cenderung mengalami "keterlambatan respons" (*lagging*) karena mencoba menghaluskan lonjakan mendadak tersebut menggunakan memori dari bulan-bulan sebelumnya.

Meskipun demikian, baik model LSTM maupun XGBoost mengalami kegagalan identik pada bulan ke-12. Secara faktual, bulan Desember didominasi oleh *Nyeri_Demam*, namun kedua algoritma kecerdasan buatan ini sangat yakin memprediksi *Saluran_Pernapasan*. Kesalahan ganda ini membuktikan bahwa terdapat anomali eksternal di luar jangkauan komputasi

murni. Secara logistik, bulan Desember merupakan periode tutup buku anggaran di mana banyak fasilitas kesehatan dan apotek melakukan penumpukan stok (*stockpiling*) untuk obat-obatan dasar seperti Paracetamol. Lonjakan angka distribusi di bulan ke-12 ini kemungkinan besar merepresentasikan pergerakan administratif rantai pasok logistik, bukan representasi epidemiologis klinis yang murni.

4.6.2 Signifikansi Model Terhadap Sistem Peringatan Dini Kesehatan

Hasil komparasi dari kedua pendekatan baik LSTM (83.33%) maupun XGBoost (91.67%) yang secara mutlak membuktikan Hipotesis Alternatif (H_1) dalam penelitian ini. Angka validasi klasifikasi, yang diukur secara presisi melalui matriks *Precision*, *Recall*, dan *F1-Score*, berada jauh di atas probabilitas acak (*random guessing*). Hal ini memberikan justifikasi ilmiah bahwa data historis distribusi obat memiliki korelasi prediktif yang valid, stabil, dan signifikan terhadap dinamika kondisi kesehatan masyarakat di wilayah Banda Aceh.

Melalui penerapan algoritma pembelajaran mesin, penelitian ini berhasil mendemonstrasikan transformasi fungsi rekam data distributor farmasi. Kumpulan *dataset* pasif yang selama ini murni digunakan sebagai dokumen penagihan dan inventaris logistik, kini berhasil direkayasa menjadi instrumen intelijen kesehatan proaktif (*Pharmacy Surveillance Information Systems*). Model prediksi multidimensi yang dihasilkan dapat diimplementasikan sebagai Sistem Pendukung Keputusan (*Decision Support System*) oleh manajemen distributor untuk memitigasi risiko kelangkaan pasokan. Di sisi makro, sistem peringatan dini (*early warning system*) ini memungkinkan pemangku kebijakan kesehatan publik untuk mendeteksi kecenderungan penyakit wabah secara tanggap, mendahului rilis data pelaporan rekam medis konvensional.

BAB V

KESIMPULAN

5.1 Kesimpulan

Berpijak pada rumusan masalah, hasil pengujian algoritmik, serta analisis komparatif yang telah diuraikan pada bab-bab sebelumnya, penelitian ini menghasilkan kesimpulan sebagai berikut:

1. Model prediksi kecenderungan penyakit berhasil dibangun secara komputasional menggunakan arsitektur *Long Short-Term Memory* (LSTM) dengan pendekatan klasifikasi deret waktu multivariat. Transformasi data logistik mentah menjadi matriks proporsi probabilitas menggunakan metode *sliding window* (memori historis 4 minggu) terbukti efektif dalam memetakan dependensi pola tren dominasi penyakit, sekaligus mereduksi bias dari lonjakan volume kuantitas absolut.
2. Hasil evaluasi metrik klasifikasi menunjukkan bahwa model LSTM memiliki keandalan yang baik dengan tingkat Akurasi sebesar 83.33% dan rata-rata *F1-Score* sebesar 0.83. Namun, analisis komparasi silang dengan algoritma *eXtreme Gradient Boosting* (XGBoost) yang berhasil mencapai Akurasi 91.67% menyingkap batasan teknis dari LSTM. Jaringan memori LSTM memiliki kecenderungan mengalami keterlambatan respons (*lagging*) dalam menangkap pergeseran wabah yang bersifat *non-linear* dan mendadak.
3. Kegagalan identik pada kedua arsitektur algoritma dalam memprediksi anomali di bulan ke-12 membuktikan bahwa data rekam distribusi farmasi masih memuat galat eksternal berupa manuver administratif (*stockpiling* atau penumpukan stok akhir tahun fiskal). Terlepas dari anomali administratif tersebut, nilai akurasi di atas 80% membuktikan secara konklusif bahwa pola historis transaksi obat di tingkat distributor merupakan indikator peringatan dini (*early warning system*) yang valid dan signifikan terhadap dinamika kesehatan publik.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang diharapkan dapat menjadi masukan bagi pengembangan penelitian selanjutnya.

Penelitian berikutnya disarankan menggunakan data historis dengan periode pengamatan yang lebih panjang sehingga model dapat mempelajari pola temporal secara lebih optimal. Dengan jumlah data yang lebih banyak, diharapkan kemampuan model dalam mengenali pola perubahan kebutuhan obat dapat meningkat sehingga menghasilkan prediksi yang lebih akurat.

Selain itu, penelitian selanjutnya dapat mempertimbangkan penambahan variabel yang berpotensi memengaruhi kebutuhan obat, seperti faktor musim, kondisi cuaca, tren penyakit, maupun data epidemiologi. Penambahan variabel tersebut diharapkan mampu meningkatkan kualitas informasi yang digunakan dalam proses pembelajaran model sehingga performa prediksi menjadi lebih baik.

Pengembangan penelitian juga dapat dilakukan melalui penyesuaian parameter maupun arsitektur model Long Short-Term Memory (LSTM) agar diperoleh konfigurasi yang lebih optimal sesuai dengan karakteristik data yang digunakan. Di samping itu, penelitian selanjutnya dapat melakukan pengujian terhadap metode prediksi lain sebagai model pembandingan untuk memperoleh gambaran mengenai performa masing-masing metode pada dataset yang berbeda.

Terakhir, hasil penelitian ini diharapkan dapat dikembangkan menjadi sistem prediksi yang terintegrasi sehingga dapat dimanfaatkan oleh distributor farmasi maupun apotek sebagai pendukung dalam perencanaan kebutuhan obat dan pengambilan keputusan berdasarkan kecenderungan penyakit.

DAFTAR PUSTAKA

- Abbasi, R., Alavi, N. M., Farzandipour, M., Gong, Y., & Nabovati, E. (2023). Using pharmacy surveillance information systems to monitor the dispensing practice of under-controlled drugs: A qualitative study on necessities, requirements, and implementation challenges. *Informatics in Medicine Unlocked*, 38(December 2022), 101198. <https://doi.org/10.1016/j.imu.2023.101198>
- Abboud, D. M., Ghazi, R., Brunaldi, V., Gala, K., Baroud, S., Kerbage, A., AbdulRazzak, F., Al Annan, K., Rapaka, B., Yao, R., Vargas, E. J., Storm, A. C., & Abu Dayyeh, B. K. (2023). Tubularization of the gastric pouch helps sustain weight loss after transoral outlet reduction for post-Roux-en-Y gastric bypass weight recurrence. *Endoscopy International Open*, 11(09), E829–E834. <https://doi.org/10.1055/a-2117-8113>
- Ankita, M., Srinivas, A., Soni, A., Prajapati, G., & Manjunath, P. S. (2024). INTELLIGENT SYSTEMS AND APPLICATIONS IN Pharmaceutical Sales Data Prediction Using Time Series Forecasting. 12, 681–696.
- Batko, K., & Ślęzak, A. (2022). The use of Big Data Analytics in healthcare. *Journal of Big Data*. <https://doi.org/10.1186/s40537-021-00553-4>
- Bontempi, G., Taieb, S. Ben, & Borgne, L. (2013). *for Time Series Forecasting*. 62–63.
- Chen, X., Lu, G., Zhang, H., Wan, J., Chen, X., Lu, G., Zhang, H., & Wan, J. (2025). Pharmaceutical Demand Forecasting via GCN-LSTM: A Knowledge Graph- Based Approach Pharmaceutical Demand Forecasting via GCN-LSTM: A Knowledge Graph-Based Approach. 0–15. <https://doi.org/10.20944/preprints202510.1546.v1>
- Daruhadi, G., & Sopiati, P. (2024). *RESEARCH DATA COLLECTION*. 04(07), 1–18.

- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). *Comparison of Convenience Sampling and Purposive Sampling*. 5(1), 1–4. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Fourkiotis, K. P., & Tsadiras, A. (2024). Applying Machine Learning and Statistical Forecasting Methods for Enhancing Pharmaceutical Sales Predictions. *Forecasting*, 6(1), 170–186. <https://doi.org/10.3390/forecast6010010>
- Hall, T., & Rasheed, K. (2025). A Survey of Machine Learning Methods for Time Series Prediction. *DL*.
- Hapsari, I. (2024). Cluster-specific Bi-LSTM models for improved pharmaceutical sales forecasting. *Journal Industrial Servicess*, 10(2), 259. <https://doi.org/10.62870/jiss.v10i2.26898>
- Heta Hemang Shah. (2023). Early Disease Detection through Data Analytics: Turning Healthcare Intelligence. 2(4), 252–269. <https://doi.org/10.47709/ijmdsa>
- Islam, M. P., & Jakarta, U. P. (2024). *Pengumpulan Data Penelitian*. 3(5), 5423–5443.
- Karimi, A., Esrafilian, A., Carty, C. P., Nasser, A., Barzan, M., Korhonen, R. K., Astori, I., Hall, W., & John, D. (2024). Computer Methods and Programs in Biomedicine Sensitivity analysis of paediatric knee kinematics to the graft surgical parameters during anterior cruciate ligament reconstruction : A sequentially linked neuromusculoskeletal-finite element analysis. *Computer Methods and Programs in Biomedicine*, 248(February), 108132. <https://doi.org/10.1016/j.cmpb.2024.108132>
- Kurniawati, A., Ahmad, M. S., Fhadli, M., & Lutfi, S. (2023). Analisis Perbandingan Metode Time Series Forecasting Untuk Prediksi Penjualan Obat Di Distributor (Studi Kasus : Kimia Farma Distributor Takoma) Comparative Analysis of Methodstime Series Forecasting for Prediction of Drug Sales in Pharmacy (Case Study : Chemica. *Jurnal Jaringan Dan ...*, 3(1), 96–106. <https://doi.org/00.0000/jati>

- Lamy, J., Sedki, K., & Tsopra, R. (2020). Explainable decision support through the learning and visualization of preferences from a formal ontology of antibiotic treatments. *Journal of Biomedical Informatics*, 104(March), 103407. <https://doi.org/10.1016/j.jbi.2020.103407>
- Mbonyinshuti, F., Nkurunziza, J., Council, N., Ncst, T., & Box, P. O. (2024). Health supply chain forecasting : a comparison of ARIMA and LSTM time series models for demand prediction of medicines. 269–280.
- Mendoza-Silva, G. M., Matey-Sanz, M., Torres-Sospedra, J., & Huerta, J. (2019). BLE RSS measurements dataset for research on accurate indoor positioning. *Data*, 4(1). <https://doi.org/10.3390/data4010012>
- Metode, M., & Ensen, A. (2023). Perancangan Sistem Informasi Penjualan Obat Berbasis Web. *Jurnal Mahasiswa Aplikasi Teknologi Komputer Dan Informasi*, 5(2), 116–120.
- Mousa, B. A., & Al-khateeb, B. (2020). *PREDICTING MEDICINE DEMAND USING DEEP*.
- Networks, S. M. (2025). G-Tech : Jurnal Teknologi Terapan Forecasting Analysis of Drug Use in Hospitals Based on Multivariate Long. 9(4), 2248–2258.
- Property, I. (2015). ICMJE Form for Disclosure of Potential Conflicts of Interest Identifying information . The work under consideration for publication . ICMJE Form for Disclosure of Potential Conflicts of Interest. *The New England Journal of Medicine*, 1–3.
- Shahzad, K., Khan, S. A., Latif, M., Muhammed, A., & Javeed, D. (2025). Big data analytics in healthcare: current practices , innovations , and future prospects.
- Supriatna, E., & Permata, S. D. (2022). Rancang Bangun Sistem Informasi Distributor Berbasis Website di Distributor Qta Farma Sumedang. *Jurnal Dimamu*, 2(1), 122–128. <https://doi.org/10.32627/dimamu.v2i1.675>
- Wardhani, Y., Hutajulu, S. H., Ferianti, V. W., Fitriani, Z., Taroeno-Hariadi, K. W., & Kurnianda, J. (2019). Effects of oxaliplatin-containing adjuvant

chemotherapy on short-term survival of patients with colon cancer in Dr. Sardjito Hospital, Yogyakarta, Indonesia. *Journal of Gastrointestinal Oncology*, 10(2), 226–234. <https://doi.org/10.21037/jgo.2018.12.10>

Wulandari, L. P. L., Sawitri, A. A. S., & Hermansyah, A. (2022). The potential roles of pharmacy medication sales data to augment the syndromic surveillance system in response to COVID-19 and preparedness for other future infectious disease outbreaks in Indonesia. *International Journal of Health Planning and Management*, 37(1), 30–39. <https://doi.org/10.1002/hpm.3356>

Yusuf, M., Basyar, S., Hayyat, M. A. M., & Rahman, F. I. (2025). Prediksi Penjualan Obat Menggunakan Model Lstm Dan Analisis Time Series Pada Data Transaksi Pasien BPJS. 7, 29–36.

