

Pendekatan Augmentasi Data Time Series Untuk Peramalan Kebutuhan Obat Antiretroviral Berbasis Long Short-Term Memory

Munazar*, Baihaqi Siregar, Amalia

Sains Data dan Kecerdasan Buatan, Ilmu Komputer, Universitas Sumatera Utara, Medan, Indonesia

Email: ^{1,*}munazarbunda@gmail.com, ²baihaqi@usu.ac.id, ³amalia@usu.ac.id

Email Penulis Korespondensi: munazarbunda@gmail.com*

Submitted: 29/05/2026; Accepted: 24/06/2026; Published: 30/06/2026

Abstrak – Perencanaan kebutuhan obat antiretroviral yang akurat diperlukan untuk menjaga kesinambungan terapi HIV, mencegah stock-out, dan meningkatkan efisiensi distribusi obat. Namun, peramalan masih terhambat oleh keterbatasan data historis, dominasi nilai nol, dan fluktuasi permintaan yang menyulitkan model mempelajari pola time series secara stabil. Penelitian ini mengevaluasi enam metode augmentasi data time series - Gaussian noise jittering, time warping, Fast Fourier Transform-based augmentation, MixUp time series, TimeGAN, dan Seasonal-Trend Decomposition using Loess — untuk meningkatkan akurasi peramalan berbasis Long Short-Term Memory. Dua skema diuji: penambahan data sintetis (25%, 50%, 75%, 100%) dan penggantian data asli (10%, 20%, 40%, 60%, 80%), dievaluasi menggunakan Mean Absolute Error non-zero. Hasil menunjukkan bahwa augmentasi moderat (25%–50%) meningkatkan akurasi, skema penambahan lebih stabil daripada penggantian, dan model terbaik menggunakan Seasonal-Trend Decomposition using Loess mencapai MAE non-zero sekitar 66.

Kata Kunci: antiretroviral; augmentasi data; forecasting; LSTM; MAE non-zero; time series.

Abstract – Accurate forecasting of antiretroviral drug demand is essential to maintain HIV treatment continuity, prevent stock-outs, and improve distribution efficiency. However, forecasting remains challenging due to limited historical data, zero-dominated observations, and fluctuating demand patterns. This study evaluates six time-series data augmentation methods — Gaussian noise jittering, time warping, Fast Fourier Transform-based augmentation, MixUp time series, TimeGAN, and Seasonal-Trend Decomposition using Loess — to improve Long Short-Term Memory-based forecasting. Two augmentation schemes were tested: synthetic data addition (25%, 50%, 75%, 100%) and original data replacement (10%, 20%, 40%, 60%, 80%), evaluated using non-zero Mean Absolute Error. Results show that moderate augmentation (25%–50%) improves accuracy, the addition scheme is more stable than replacement, and the best model using Seasonal-Trend Decomposition using Loess achieved a non-zero MAE of approximately 66.

Keywords: antiretroviral drugs; data augmentation; forecasting; Long Short-Term Memory; time series.

1. PENDAHULUAN

Ketersediaan obat antiretroviral merupakan aspek krusial dalam keberlanjutan terapi bagi orang dengan HIV. Kesalahan estimasi kebutuhan obat menimbulkan dua risiko utama: kekosongan stok (*stock-out*) yang mengganggu kesinambungan terapi, dan kelebihan stok (*overstock*) yang menyebabkan pemborosan anggaran serta risiko kedaluwarsa obat. Di Indonesia, tantangan ini semakin kompleks karena kebutuhan obat bervariasi antarwilayah, dipengaruhi jumlah pasien aktif, pola *regimen*, dan dinamika distribusi logistik [23]. Oleh karena itu, diperlukan pendekatan *forecasting* berbasis data historis yang mampu memprediksi kebutuhan obat secara lebih akurat.

Permasalahan utama dalam penelitian ini adalah keterbatasan data historis penggunaan obat antiretroviral. Data *time series* kebutuhan obat umumnya bersifat tidak seimbang, didominasi nilai nol, dan memiliki fluktuasi permintaan yang tidak konsisten [9]. Ketidakteraturan ini menyulitkan model *machine learning* mempelajari pola temporal secara optimal [26]. Khususnya pada model deep learning seperti Long Short-Term Memory (LSTM), data yang terbatas berisiko menyebabkan *overfitting* sehingga model gagal melakukan generalisasi pada data baru [1].

Salah satu solusi yang ditawarkan dalam penelitian ini adalah penerapan data *augmentation* pada data *time series* penggunaan obat antiretroviral. Data *augmentation* merupakan teknik untuk memperkaya data pelatihan dengan membentuk variasi data baru dari data asli tanpa harus melakukan pengumpulan data baru secara langsung. Dalam penelitian ini, augmentasi data tidak dipahami hanya sebagai proses menambah jumlah data, tetapi sebagai pendekatan rekayasa data yang harus mempertahankan distribusi, pola temporal, dan struktur statistik data asli. Oleh karena itu, penelitian ini mengevaluasi beberapa teknik augmentasi numerik, yaitu *Gaussian noise jittering*, *time warping*, *Fast Fourier Transform-based augmentation*, *MixUp time series*, *TimeGAN*, dan *Seasonal-Trend Decomposition using Loess*. Seluruh teknik tersebut diuji menggunakan model Long Short-Term Memory untuk mengetahui pengaruhnya terhadap akurasi *forecasting* kebutuhan obat antiretroviral.

Sejumlah penelitian terkait menunjukkan potensi augmentasi data untuk meningkatkan performa model pada kondisi data terbatas. Dari sisi teknik augmentasi time series secara umum, Wen et al. [14] dan Iglesias et al. [3] menegaskan bahwa pemilihan metode yang sesuai dengan karakteristik data merupakan faktor kunci keberhasilan augmentasi untuk deep learning. Gao et al. [4] membuktikan bahwa tidak semua teknik memberikan

peningkatan akurasi secara konsisten, bergantung pada jenis dataset dan arsitektur model yang digunakan. Stenger et al. [6] menekankan bahwa kualitas data sintetis harus dinilai dari kontribusinya terhadap performa model prediktif, bukan sekadar kuantitasnya. Terkait metode spesifik yang diuji dalam penelitian ini, Yoon et al. [19] mengembangkan TimeGAN yang mampu menghasilkan data sintetis time series dengan mempertahankan dependensi temporal. Hao dan Liu [21] serta Liang et al. [24] menunjukkan bahwa dekomposisi berbasis STL secara konsisten meningkatkan akurasi peramalan LSTM pada data dengan pola musiman dan tren yang kompleks. Alexander [22] mempertegas bahwa metode berbasis dekomposisi dan Gaussian noise memberikan peningkatan akurasi yang lebih stabil dibandingkan metode augmentasi lainnya. Dalam konteks peramalan kebutuhan obat, Pall et al. [18] membuktikan bahwa model machine learning dapat memprediksi kekurangan obat secara proaktif, sedangkan Morid et al. [20] menegaskan bahwa LSTM pada data farmasi yang terbatas rentan terhadap overfitting sehingga strategi pengayaan data menjadi kritis. Li et al. [25] memperkuat argumen ini dengan menunjukkan bahwa komposisi yang seimbang antara data asli dan sintetis merupakan faktor penentu utama stabilitas dan akurasi model forecasting.

Berdasarkan kajian di atas, terdapat celah penelitian yang belum banyak dieksplorasi. Sebagian besar studi terdahulu berfokus pada klasifikasi time series atau domain umum, bukan pada forecasting kebutuhan obat dengan data yang sangat terbatas dan sparse. Selain itu, pengaruh komposisi data sintetis baik melalui skema penambahan maupun penggantian — terhadap performa LSTM belum dikaji secara sistematis. Padahal, proporsi data sintetis yang tidak tepat berpotensi menyebabkan distribution shift yang menurunkan kemampuan generalisasi model.

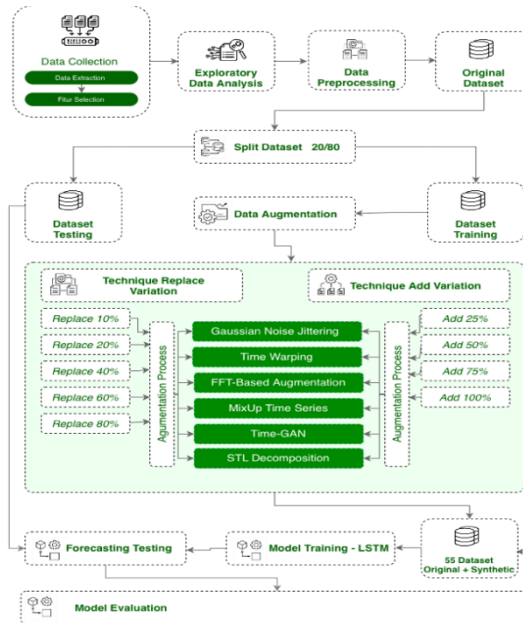
Kebaruan penelitian ini terletak pada evaluasi sistematis enam teknik augmentasi data time series Gaussian noise jittering, time warping, FFT-based augmentation, MixUp, TimeGAN, dan STL decomposition untuk meningkatkan akurasi forecasting kebutuhan obat antiretroviral menggunakan LSTM. Penelitian ini juga menguji dua strategi komposisi data sintetis: add variation (25%, 50%, 75%, 100%) dan replace variation (10%, 20%, 40%, 60%, 80%). Tujuan penelitian ini adalah menentukan metode dan komposisi augmentasi yang paling efektif untuk menjaga pola temporal dan meningkatkan akurasi prediksi, sehingga hasil penelitian dapat memberikan rekomendasi praktis bagi perencanaan stok obat antiretroviral yang lebih akurat dan berbasis data.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan utama untuk mengevaluasi efektivitas teknik augmentasi data time series dalam meningkatkan akurasi forecasting kebutuhan obat antiretroviral menggunakan model Long Short-Term Memory. Tahapan penelitian dimulai dari pengumpulan data, ekstraksi data, seleksi fitur, eksplorasi data, preprocessing, pembagian dataset, proses augmentasi data, normalisasi, pelatihan model, evaluasi model, hingga analisis hasil. Seluruh tahapan dirancang untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik, metode augmentasi diterapkan secara sistematis, dan hasil pengujian dapat dibandingkan secara objektif. Metodologi ini disusun berdasarkan rancangan penelitian pada tesis yang telah dikembangkan sebelumnya.

Secara umum, alur tahapan penelitian dapat dijelaskan sebagai berikut. Pertama, data penggunaan obat antiretroviral dikumpulkan dari sistem forecasting yang digunakan dalam proses distribusi obat antiretroviral di Indonesia. Kedua, dilakukan ekstraksi data dari beberapa tabel basis data, kemudian data digabungkan untuk memperoleh informasi yang relevan. Ketiga, dilakukan seleksi fitur untuk menghapus atribut yang tidak digunakan dalam proses pemodelan. Keempat, dilakukan eksplorasi data untuk memahami struktur, tipe data, nilai kosong, serta karakteristik awal dataset. Kelima, dilakukan preprocessing untuk menangani nilai kosong, mengubah data kategorikal menjadi numerik, dan menyiapkan data agar dapat digunakan oleh model. Keenam, dataset dibagi menjadi data latih dan data uji. Ketujuh, data latih diproses menggunakan beberapa metode augmentasi dengan skema komposisi berbeda. Kedelapan, hasil augmentasi dinormalisasi menggunakan Min-Max Scaling. Kesembilan, model Long Short-Term Memory dilatih menggunakan dataset asli dan dataset hasil augmentasi. Terakhir, performa model dievaluasi menggunakan Mean Absolute Error pada observasi non-zero.



Gambar 1. Arsitektur Umum

Gambar 1 menunjukkan alur penelitian yang digunakan untuk menguji pengaruh augmentasi data terhadap akurasi forecasting. Tahapan tersebut menggambarkan bahwa augmentasi hanya diterapkan pada data latih, sedangkan data uji tetap menggunakan data asli agar evaluasi model tetap objektif dan tidak terjadi data leakage.

Sebelum di lakukan proses augmentasi dataset dibagi menjadi dua bagian dengan komposisi 20% digunakan untuk data testing dan 80% digunakan sebagai data training.

2.2 Dataset Penelitian

Dataset yang digunakan dalam penelitian ini merupakan data historis penggunaan obat antiretroviral dari berbagai provinsi di Indonesia. Data diperoleh dari sistem yang digunakan untuk mendukung proses distribusi obat antiretroviral. Proses ekstraksi dilakukan dari basis data MySQL dengan mengambil tabel-tabel yang memuat informasi layanan kesehatan, rejimen obat, dan riwayat penyaluran obat. Data yang telah diekstraksi kemudian dikonversi ke format .csv agar lebih mudah diproses menggunakan bahasa pemrograman Python.

Tabel 1 menunjukkan sumber data yang digunakan dalam penelitian.

Tabel 1. Dataset Sumber Penelitian

No.	Tabel	Keterangan	Jumlah Baris	Jumlah Kolom
1.	ref_layanan	Menyimpan informasi data master layanan kesehatan sumber data penyaluran obat	14.131	19
2.	ref_rejimen	Menyimpan informasi master daftar nama-nama rejimen obat ARV	592	10
3.	trans_historical_data	Menyimpan data riwayat penyaluran obat ARV	2.239.222	19

2.3 Pengecekan Struktur dan Tipe Data

Data yang sudah di muat kedalam variable menggunakan library panda di gunakan perintah .info() untuk melihat jumlah baris data, jumlah kolom dan tipe data setiap kolom ditampilkan pada gambar dibawah ini :

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2239222 entries, 0 to 2239221
Data columns (total 11 columns):
#   Column      Dtype
---  ---
0   tahun       int64
1   bulan       int64
2   jml_hari    int64
3   rejimen_id  int64
4   prov_id     int64
5   kab_id     int64
6   layanan_id int64
7   jml_pasien int64
8   kategori    object
9   kode        object
10  nama        object
dtypes: int64(8), object(3)
memory usage: 187.9+ MB
```

Gambar 2. Struktur dan Tipe Data

2.4 Data Preprocessing

Pada tahap ini dilakukan pembersihan dataset terhadap beberapa fitur dan tipe data yang sudah difahami pada tahapan EDA. Tahapan ini dilakukan sebelum dilakukan splitting dataset untuk kebutuhan training model machine learning. Tujuannya untuk memilah data yang memiliki format yang tidak sesuai dengan input yang dibutuhkan oleh model, sehingga optimasi algoritma model augmentasi dan forecasting yang digunakan menjadi stabil. Kolom kode dan nama akan di hapus dari dataset, dan kolom kategori dilakukan label encoding untuk mengubah isi kategorial A dan D menggunakan library sklearn.preprocessing

Tabel 2. Data Hasil Preprosesing

	tahun	bulan	jml_hari	rejimen_id	prov_id	kab_id	layanan_id	jml_pasien	kategori_encoded
0	2021	1	31	1	11	1173	1105012	4	1
1	2021	1	31	1	11	1173	1105012	0	0
2	2021	1	31	2	11	1173	1105012	5	1
3	2021	1	31	2	11	1173	1105012	0	0
4	2021	1	31	3	11	1173	1105012	8	1
...
2239217	2024	8	31	322	76	7606	7606050101	0	0
2239218	2024	8	31	517	76	7606	94141090691	0	1
2239219	2024	8	31	517	76	7606	94141090691	0	0
2239220	2024	8	31	14	76	7606	94141090692	0	1
2239221	2024	8	31	14	76	7606	94141090692	0	0

Dari tabel 2 dapat disimpulkan bahwa dataset sudah sangat normal untuk dapat digunakan pada proses augmentasi dan training model forecasting karena sudah tidak ada lagi fitur dengan tipe data object.

2.5 Metode Augmentasi Data

Augmentasi data dilakukan untuk memperkaya variasi data latih tanpa mengubah data uji. Penelitian ini menggunakan enam metode augmentasi data numerik yang mewakili beberapa kategori teknik augmentasi time series, yaitu random transformation, time-domain augmentation, frequency-domain augmentation, pattern mixing, generative model, dan decomposition-based augmentation.

Metode pertama adalah Gaussian noise jittering, yaitu teknik augmentasi yang menambahkan gangguan acak berdistribusi normal pada data asli. Metode ini bertujuan menghasilkan variasi kecil yang menyerupai fluktuasi alami. Metode kedua adalah time warping, yaitu teknik yang memodifikasi sumbu waktu dengan cara meregangkan atau memampatkan pola tertentu pada deret waktu. Metode ketiga adalah Fast Fourier Transform-based augmentation, yaitu teknik yang memanfaatkan transformasi frekuensi untuk menghasilkan variasi data dengan tetap mempertahankan pola periodik. Metode keempat adalah MixUp time series, yaitu teknik yang menghasilkan data sintetis melalui pencampuran linear antara dua sampel deret waktu, yang pertama kali diperkenalkan oleh Zhang et al. [8] sebagai strategi regularisasi berbasis interpolasi untuk meningkatkan generalisasi model. Metode kelima adalah TimeGAN, yaitu model generatif yang dirancang untuk menghasilkan data sintetis dengan mempertahankan dependensi temporal [10]. TimeGAN dibangun di atas kerangka Generative Adversarial Network yang dikembangkan oleh Goodfellow et al. [11] dan konsep variational autoencoder dari Kingma dan Welling [12], serta terinspirasi dari penelitian Esteban et al. [16] yang menerapkan GAN rekuren untuk menghasilkan data medis time series yang realistis. Metode keenam adalah Seasonal-Trend Decomposition using Loess, yaitu teknik dekomposisi yang memisahkan deret waktu menjadi komponen tren, musiman, dan residual sebelum dilakukan pembentukan variasi data.

Tabel 3. Metode Augmentasi Data

No	Kategori	Metode	Tujuan
1.	Random transformation	Gaussian noise jittering	Menambahkan variasi kecil pada data asli
2.	Time-domain augmentation	Time warping	Mengubah dinamika waktu tanpa menghilangkan pola utama
3.	Frequency-domain augmentation	Transform-based augmentation	Mempertahankan pola periodik dan variasi frekuensi

No	Kategori	Metode	Tujuan
4.	Pattern mixing	MixUp time series	Menggabungkan dua pola deret waktu
5.	Generative model	TimeGAN	Menghasilkan data sintetis berbasis pola temporal
6.	Decomposition-based augmentation	Seasonal-Trend Decomposition using Loess	Membentuk variasi dari komponen tren, musiman, dan residual

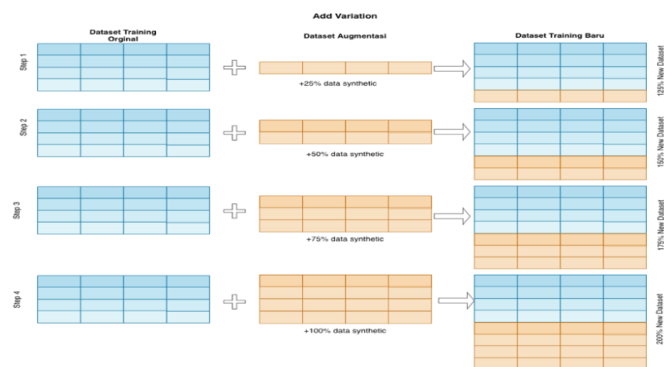
2.6 Skenario Eksperimen

Penelitian ini menggunakan dua skenario utama augmentasi, yaitu *add variation* dan *replace variation*. Pada skenario *add variation*, data sintetis ditambahkan ke data asli sehingga ukuran dataset latih meningkat. Komposisi data sintetis yang digunakan adalah 25%, 50%, 75%, dan 100% dari jumlah data latih. Skenario ini bertujuan untuk mengetahui sejauh mana penambahan data sintetis dapat meningkatkan kemampuan model dalam mempelajari pola temporal.

Pada skenario *replace variation*, sebagian data asli diganti dengan data sintetis tanpa mengubah ukuran dataset latih. Komposisi penggantian data yang digunakan adalah 10%, 20%, 40%, 60%, dan 80%. Skenario ini bertujuan untuk menguji dampak penggantian data asli terhadap stabilitas distribusi data dan akurasi prediksi model.

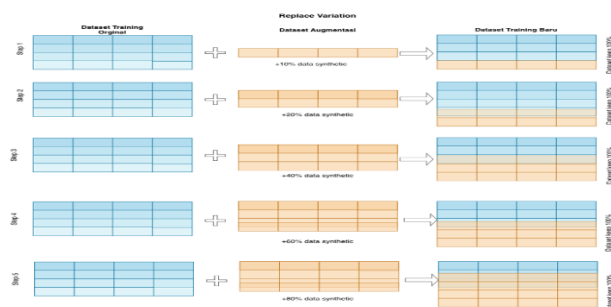
Tabel 4. Skenario Komposisi Augmentasi

No	Skenario	Komposisi Data Sintetis	Keterangan
1.	<i>Add variation</i>	25%, 50%, 75%, 100%	Data sintetis ditambahkan ke data asli
2.	<i>Replace variation</i>	10%, 20%, 40%, 60%, 80%	Sebagian data asli diganti dengan data sintetis



Gambar 3 Data Sintetis *Add Variant*

Gambar 3 menunjukkan skema *add variation*, yaitu proses augmentasi dengan menambahkan data sintetis ke dalam data asli tanpa menghapus data awal. Pada skema ini, jumlah data latih akan meningkat sesuai dengan proporsi augmentasi yang digunakan, yaitu 25%, 50%, 75%, dan 100%. Pendekatan ini bertujuan untuk memperkaya variasi data pelatihan sehingga model dapat mempelajari pola temporal dengan lebih baik, namun tetap mempertahankan distribusi data asli.



Gambar 4. Data Sintetis *Replace Variant*

Gambar 4 menunjukkan skema replace variation, yaitu proses augmentasi dengan mengganti sebagian data asli menggunakan data sintetis dalam proporsi tertentu. Berbeda dengan add variation, pada skema ini ukuran dataset tetap sama, tetapi komposisi antara data asli dan data sintetis berubah sesuai tingkat penggantian, yaitu 10%, 20%, 40%, 60%, dan 80%. Pendekatan ini digunakan untuk mengevaluasi sejauh mana data sintetis dapat menggantikan data asli tanpa mengurangi kemampuan model dalam mempertahankan pola temporal dan akurasi prediksi.

2.7 Model Long Short-Term Memory

Model yang digunakan dalam penelitian ini adalah Long Short-Term Memory. Model ini dipilih karena mampu mempelajari dependensi jangka panjang pada data time series. Berbeda dengan jaringan saraf biasa, Long Short-Term Memory memiliki mekanisme gerbang yang dapat mengatur informasi mana yang perlu disimpan, diperbarui, atau dilupakan. Kemampuan ini membuat model sesuai digunakan untuk peramalan kebutuhan obat antiretroviral yang memiliki pola temporal, fluktuasi, dan variasi permintaan.

Pada penelitian ini, model dilatih menggunakan beberapa dataset, yaitu dataset asli dan dataset hasil augmentasi dari setiap metode serta skenario komposisi. Dengan demikian, setiap konfigurasi augmentasi menghasilkan satu model yang kemudian dibandingkan berdasarkan nilai kesalahan prediksi. Perbandingan ini digunakan untuk menentukan metode dan komposisi augmentasi yang paling efektif dalam meningkatkan akurasi forecasting.

2.8 Evaluasi Model Forecasting

Evaluasi model dilakukan menggunakan Mean Absolute Error pada observasi non-zero. Metrik ini digunakan karena data penggunaan obat antiretroviral memiliki karakteristik sparse atau banyak mengandung nilai nol. Jika evaluasi dilakukan pada seluruh observasi, nilai nol yang dominan dapat membuat hasil evaluasi terlihat lebih baik, tetapi tidak merepresentasikan kemampuan model dalam memprediksi kondisi ketika terdapat kebutuhan obat. Oleh karena itu, pengukuran difokuskan pada observasi dengan nilai aktual bukan nol.

Rumus MAE_{non zero} dinyatakan sebagai berikut:

$$MAE_{non\ zero} = \frac{1}{m} \sum_{i \in S} |y_i - \hat{y}_i| \tag{1}$$

Dengan keterangan:

$S = \{i \mid y_i \neq 0\}$ adalah himpunan observasi dengan nilai aktual non-zero,

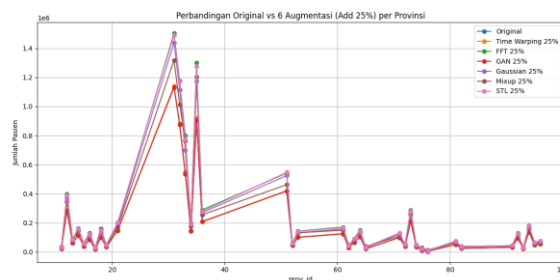
m adalah jumlah observasi dalam himpunan tersebut,

y_i dan \hat{y}_i masing-masing adalah nilai aktual dan prediksi.

3. HASIL DAN PEMBAHASAN

3.1 Perbandingan Hasil Augmentasi Data

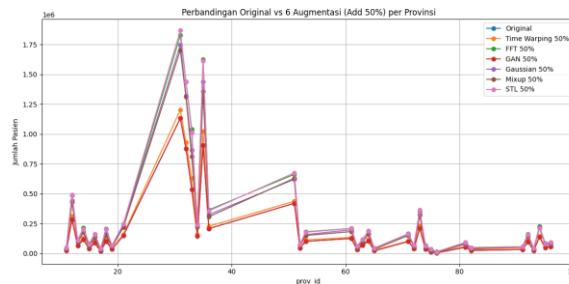
Hasil ini menunjukkan bahwa augmentasi data tidak dapat hanya dilihat sebagai proses memperbanyak data, tetapi harus mempertimbangkan keseimbangan antara data asli dan data sintetis. Pada data numerik dan time series, kualitas data sintetis sangat menentukan keberhasilan augmentasi. Data sintetis yang terlalu banyak atau tidak sesuai dengan pola asli dapat menyebabkan distribution shift, yaitu pergeseran distribusi antara data latih dan data uji. Pergeseran tersebut dapat membuat model kehilangan kemampuan dalam memprediksi data aktual secara akurat.



Gambar 5. Hasil Augmentasi Add Variant 25%

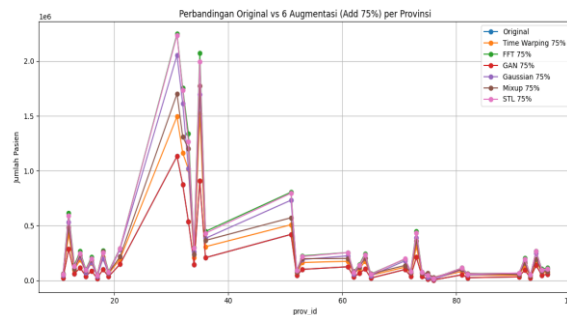
Pada skenario *add variation 25%*, seluruh metode augmentasi masih mampu mempertahankan struktur utama deret waktu asli, terutama tren dan fluktuasi utama. *Time Warping* dan *FFT-Based Augmentation* menghasilkan variasi yang relatif halus, *Time-GAN* cukup adaptif dalam meniru dinamika data, sedangkan *MixUp*

Time Series cenderung meratakan nilai ekstrem. Secara umum, distorsi pada proporsi ini masih rendah sehingga penambahan 25% data sintesis dinilai cukup aman untuk memperkaya data latih tanpa mengubah distribusi asli secara signifikan.



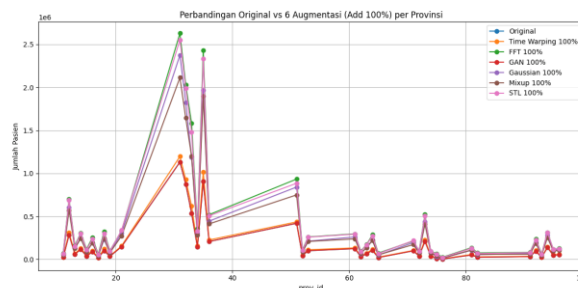
Gambar 6. Hasil Augmentasi Add Variant 50%

Pada skenario *add variation* 50%, pola tren utama masih dapat dipertahankan, tetapi perbedaan antar metode mulai terlihat lebih jelas. *Time Warping* dan *FFT-Based Augmentation* masih relatif stabil meskipun terdapat sedikit pergeseran amplitudo, sementara *Time-GAN* tetap mengikuti pola asli dengan peningkatan *noise* pada beberapa titik. *MixUp Time Series* menghasilkan pola yang lebih halus dan cenderung mengurangi fluktuasi ekstrem. Secara umum, augmentasi 50% masih mampu menjaga karakteristik data, tetapi mulai menunjukkan peningkatan distorsi dibandingkan proporsi 25%.



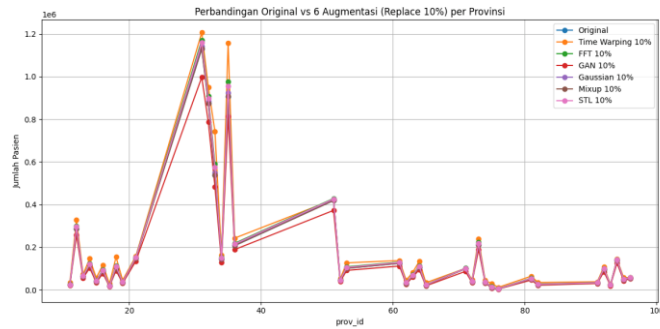
Gambar 7. Hasil Augmentasi Add Variant 75%

Pada skenario *add variation* 75%, perbedaan antar metode semakin terlihat dan distorsi terhadap pola asli mulai meningkat. Tren umum masih dapat dikenali, tetapi beberapa puncak dan fluktuasi tidak lagi sepenuhnya konsisten dengan data asli. *Time Warping* dan *FFT-Based Augmentation* menunjukkan pergeseran amplitudo yang lebih besar, *Time-GAN* masih mengikuti pola global tetapi dengan peningkatan *noise*, sedangkan *MixUp Time Series* menghasilkan pola yang lebih rata. Secara umum, augmentasi 75% menambah variasi data secara signifikan, tetapi mulai menurunkan kemiripan terhadap pola temporal asli.



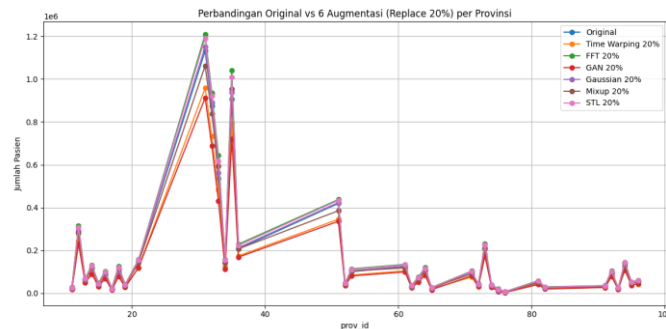
Gambar 8. Hasil Augmentasi Add Variant 100%

Pada skenario *add variation* 100%, distorsi terhadap pola asli terlihat paling besar dibandingkan proporsi sebelumnya. Tren umum masih dapat dikenali, tetapi detail penting seperti puncak dan fluktuasi tajam mulai berkurang. *Time Warping* dan *FFT-Based Augmentation* menunjukkan perubahan amplitudo yang cukup besar, *Time-GAN* menghasilkan *noise* lebih tinggi, sedangkan *MixUp Time Series* semakin meratakan pola sehingga variasi ekstrem hampir hilang. Hal ini menunjukkan bahwa penambahan data sintesis yang terlalu besar dapat menurunkan realisme data dan berpotensi mengurangi akurasi *forecasting*.



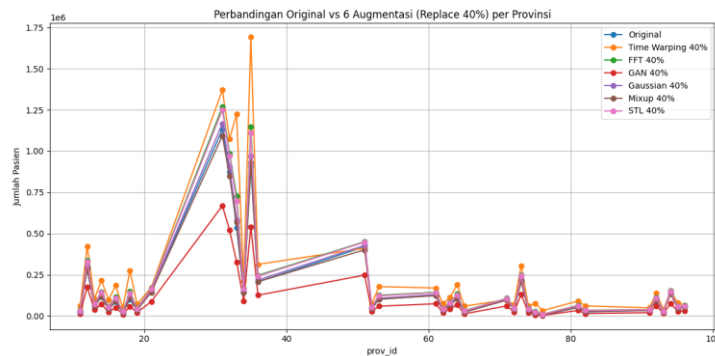
Gambar 9. Hasil Augmentasi Replace Variant 10%

Pada skenario penggantian 10% data dengan data sintetis seperti gambar 4.5 pola deret waktu masih sangat mendekati data asli. Tren, puncak, dan fluktuasi utama tetap terjaga dengan baik, serta hanya terdapat sedikit perbedaan pada beberapa titik. Seluruh metode augmentasi menunjukkan hasil yang stabil dengan tingkat noise yang rendah. Time-GAN dan FFT-Based cenderung paling mendekati pola asli, sementara Mix-Up sedikit menghaluskan variasi ekstrem. Secara keseluruhan, tingkat distorsi sangat minimal, sehingga skenario ini efektif untuk meningkatkan variasi data tanpa mengganggu karakteristik utama deret waktu.



Gambar 10. Hasil Augmentasi Hasil Replace Variant 20%

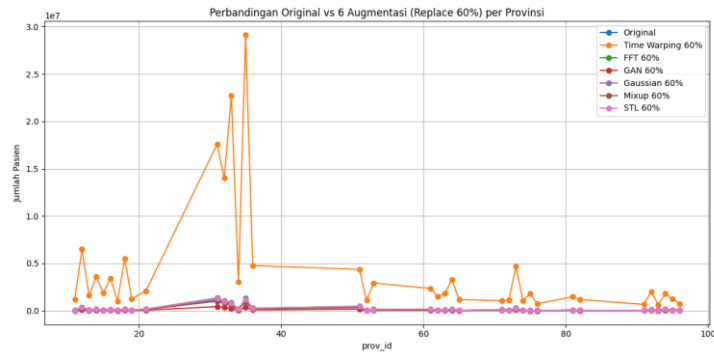
Pada penggantian 20% data dengan data sintetis seperti gambar 4.6 pola deret waktu secara umum masih dapat dipertahankan dengan baik. Tren utama dan sebagian besar puncak permintaan masih mengikuti pola asli, meskipun mulai terlihat sedikit perbedaan pada beberapa titik transisi. Variasi antar metode mulai muncul, namun belum menunjukkan penyimpangan yang signifikan. Dari sisi kualitas, noise mulai meningkat meskipun masih dalam batas wajar. Metode berbasis model seperti Time-GAN cenderung tetap adaptif, sementara metode lain menunjukkan sedikit pergeseran amplitudo. Secara keseluruhan, skenario ini masih cukup efektif karena mampu menjaga keseimbangan antara variasi data dan kemiripan terhadap pola asli.



Gambar 11. Hasil Augmentasi Replace Variant 40%

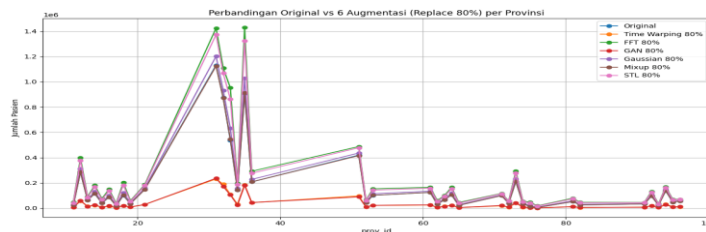
Pada level penggantian 40% pada gambar 4.7 perbedaan antara data asli dan sintetis mulai terlihat lebih jelas. Beberapa pola puncak dan fluktuasi tajam tidak lagi direpresentasikan secara konsisten oleh semua metode. Meskipun tren global masih dapat dikenali, detail lokal mulai mengalami distorsi. Selain itu, variabilitas data meningkat seiring dengan bertambahnya proporsi data sintetis. Noise dan perubahan amplitudo menjadi lebih

dominan, terutama pada metode tertentu. Hal ini menunjukkan bahwa peningkatan proporsi penggantian mulai berdampak pada penurunan kualitas representasi data.



Gambar 12. Hasil Augmentasi Replace Variant 60%

Pada penggantian 60% lihat gambar 4.8 distorsi terhadap pola asli menjadi semakin signifikan. Banyak bagian dari deret waktu tidak lagi mengikuti karakteristik awal, terutama pada segmen dengan fluktuasi tinggi. Pola musiman dan lonjakan permintaan mulai sulit dikenali secara konsisten. Dari sisi stabilitas, hampir seluruh metode menunjukkan peningkatan noise dan penyimpangan yang cukup besar. Perbedaan antar metode juga semakin mencolok, menandakan bahwa kemampuan masing-masing teknik dalam mempertahankan pola mulai terbatas. Kondisi ini mengindikasikan penurunan reliabilitas data sintesis.

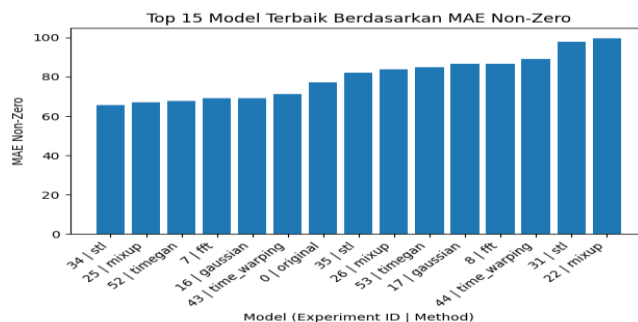


Gambar 13. Hasil Augmentasi Replace Variant 80%

Pada level 80% lihat gambar 4.9 sebagian besar pola asli telah mengalami degradasi yang cukup besar. Tren umum masih dapat terlihat, namun detail penting seperti puncak permintaan dan variasi musiman banyak yang hilang atau berubah secara signifikan. Hal ini membuat data sintesis semakin jauh dari kondisi nyata. Selain itu, tingkat noise dan distorsi menjadi sangat tinggi, sehingga konsistensi antar metode semakin rendah. Data yang dihasilkan cenderung kurang representatif dan berpotensi menurunkan performa model forecasting. Dengan demikian, proporsi penggantian sebesar ini tidak direkomendasikan karena mengurangi realisme data secara signifikan.

3.2 Hasil Forecasting

Dari semua data hasil augmentasi dan *baseline* data didapatkan model dengan kinerja terbaik yang diurutkan dari hasil evaluasi kinerja model :



Gambar 14. Grafik Hasil Evaluasi Kinerja Model *Forecasting*

Evaluasi kinerja model dilakukan menggunakan metrik *Mean Absolute Error (MAE) non-zero*, yaitu metrik yang mengukur rata-rata kesalahan prediksi hanya pada data dengan nilai aktual tidak nol. Pendekatan ini digunakan untuk menghindari bias evaluasi akibat dominasi nilai nol pada dataset, sehingga hasil evaluasi lebih merepresentasikan kemampuan model dalam memprediksi data yang benar-benar memiliki nilai informatif. Nilai *MAE non-zero* menunjukkan rata-rata deviasi absolut antara nilai prediksi dan nilai aktual; semakin kecil nilai MAE, semakin tinggi akurasi model.

Gambar 14 menunjukkan 15 model terbaik berdasarkan nilai MAE non-zero, dengan rentang nilai sekitar 65 hingga 100. Model terbaik diperoleh pada *experiment ID 34* menggunakan metode *STL decomposition* dengan nilai MAE terendah sekitar 66, yang berarti rata-rata kesalahan prediksi model sebesar 66 unit dari nilai aktual. Sementara itu, model dengan performa terendah dalam kelompok ini adalah *experiment ID 22* menggunakan metode *MixUp*, dengan MAE mendekati 100. Selisih lebih dari 30 unit antara model terbaik dan peringkat ke-15 menunjukkan adanya perbedaan performa yang cukup signifikan antar konfigurasi model.

4. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa pendekatan augmentasi data *time series* mampu menjadi solusi untuk mengatasi keterbatasan data historis dalam peramalan kebutuhan obat antiretroviral. Permasalahan utama pada penelitian ini adalah terbatasnya data pelatihan, dominasi nilai nol, serta adanya fluktuasi permintaan yang menyebabkan model sulit menangkap pola temporal secara stabil. Melalui pengujian enam metode augmentasi, yaitu *Gaussian noise jittering*, *time warping*, *Fast Fourier Transform-based augmentation*, *MixUp time series*, *TimeGAN*, dan *Seasonal-Trend Decomposition using Loess*, penelitian ini menunjukkan bahwa augmentasi data dapat meningkatkan kinerja model *Long Short-Term Memory* apabila diterapkan dengan metode dan proporsi yang tepat. Hasil pengujian menunjukkan bahwa skema *add variation* lebih stabil dibandingkan *replace variation* karena data asli tetap dipertahankan, sedangkan data sintetis hanya digunakan sebagai tambahan variasi pelatihan. Proporsi augmentasi yang moderat, khususnya pada rentang 25% hingga 50%, mampu menjaga keseimbangan antara peningkatan jumlah data dan pelestarian pola temporal asli. Sebaliknya, penggunaan data sintetis dalam proporsi terlalu besar dapat menyebabkan distorsi distribusi, peningkatan noise, dan penurunan kemiripan terhadap data aktual. Model terbaik diperoleh menggunakan metode *Seasonal-Trend Decomposition using Loess* dengan nilai *Mean Absolute Error non-zero* sekitar 66, yang menunjukkan bahwa metode berbasis dekomposisi mampu mempertahankan struktur tren, musiman, dan residual secara lebih baik dibandingkan beberapa metode lainnya. Keterbatasan penelitian ini terletak pada penggunaan satu model utama, yaitu *Long Short-Term Memory*, serta evaluasi yang masih berfokus pada satu metrik utama, yaitu *Mean Absolute Error non-zero*. Selain itu, penelitian ini belum membahas aspek implementasi langsung dalam sistem logistik obat, seperti kebijakan distribusi, kapasitas gudang, dan faktor eksternal yang dapat memengaruhi kebutuhan obat antiretroviral. Oleh karena itu, penelitian selanjutnya dapat menguji metode augmentasi pada model lain, seperti *Gated Recurrent Unit*, *Temporal Convolutional Network*, atau *Transformer-based forecasting*, serta menambahkan variabel eksternal agar hasil prediksi lebih komprehensif dan relevan untuk mendukung pengambilan keputusan dalam perencanaan stok obat antiretroviral.

REFERENCES

- [1] S. Ekwaro-Osire et al., "Numerical data augmentation techniques for machine learning: Methods and challenges," *Journal of Data Science and Artificial Intelligence*, 2025.
- [2] M. Annaki et al., "Evaluating consistency in numerical data augmentation for time series forecasting," *IEEE Access*, 2024.
- [3] Y. Tan et al., "Impact of synthetic data proportion on forecasting performance," *Expert Systems with Applications*, 2025.
- [4] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 60, 2019.
- [5] B. K. Iwana and S. Uchida, "An empirical survey of data augmentation for time series classification with neural networks," *PLOS ONE*, vol. 15, no. 7, p. e0235799, 2020.
- [6] T. T. Um et al., "Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks," in *Proc. 19th ACM Int. Conf. Multimodal Interaction*, 2017, pp. 216–220.
- [7] Q. Wen et al., "Time series data augmentation for deep learning: A survey," in *Proc. 30th Int. Joint Conf. Artificial Intelligence (IJCAI)*, 2021.
- [8] H. Zhang et al., "mixup: Beyond empirical risk minimization," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2018.
- [9] G. Forestier et al., "Generating synthetic time series to augment sparse datasets," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, 2017, pp. 865–872.
- [10] J. Yoon, D. Jarrett, and M. van der Schaar, "Time-series generative adversarial networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.
- [11] I. Goodfellow et al., "Generative adversarial nets," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.
- [12] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," *arXiv preprint arXiv:1312.6114*, 2013.

- [13] H. I. Fawaz et al., “Data augmentation using synthetic data for time series classification with deep residual networks,” *IEEE Transactions on Neural Networks and Learning Systems*, 2018.
- [14] Q. Wen et al., “Time series data augmentation: A comprehensive survey,” *ACM Computing Surveys*, 2022.
- [15] U. Schlegel et al., “A taxonomy of numerical data augmentation techniques,” *Data Mining and Knowledge Discovery*, 2024.
- [16] C. Esteban et al., “Real-valued medical time series generation with recurrent conditional GANs,” *arXiv preprint arXiv:1706.02633*, 2017.
- [17] M. A. Morig, O. R. L. Sheng, and J. Dunbar, “Time series prediction using deep learning methods in healthcare,” *ACM Transactions on Management Information Systems*, vol. 14, no. 1, pp. 1–31, 2023, doi: 10.1145/3531326.
- [18] R. Pall, Y. Gauthier, S. Auer, and W. Mowaswes, “Predicting drug shortages using pharmacy data and machine learning,” *Health Care Management Science*, vol. 26, no. 3, pp. 395–411, 2023, doi: 10.1007/s10729-022-09627-y.
- [19] J. Yoon, D. Jarrett, and M. van der Schaar, “Time-series generative adversarial networks,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, 2019, pp. 5508–5518.
- [20] A. Mak, A. Aamer, and W. Mowaswes, “Demand forecasting model for time-series pharmaceutical data using shallow and deep neural network model,” *Neural Computing and Applications*, vol. 35, no. 2, pp. 1233–1247, 2023, doi: 10.1007/s00521-022-07889-9.
- [21] J. Hao and F. Liu, “Improving long-term multivariate time series forecasting with a seasonal-trend decomposition-based 2-dimensional temporal convolution dense network,” *Scientific Reports*, vol. 14, no. 1, p. 1739, 2024, doi: 10.1038/s41598-024-52240-y.
- [22] V. Alexander, “Enhancing time series data predictions: A survey of augmentation techniques and model performances,” in *Proc. 2024 Australasian Computer Science Week (ACSW 2024)*, ACM, 2024, pp. 1–19, doi: 10.1145/3641142.3641143.
- [23] M. T. Zeleke, B. W. Atsbeha, B. Y. Melaku, Y. T. Mekasha, A. W. Mekonen, and S. D. Nigatu, “Performance of antiretroviral drugs supply chain management and related challenges in Amhara National Regional State, Ethiopia,” *Research in Clinical and Social Pharmacy*, vol. 18, p. 100570, 2025, doi: 10.1016/j.rcsop.2025.100570.
- [24] Y. Liang, J. Keylock, J. L. Peng, and L. Wu, “Effective LSTMs with seasonal-trend decomposition and adaptive learning and niching-based backtracking search algorithm for time series forecasting,” *Expert Systems with Applications*, vol. 236, p. 121275, 2024, doi: 10.1016/j.eswa.2023.121275.
- [25] X. Li, A. H. H. Ngu, and V. Metsis, “TTS-CGAN: A transformer time-series conditional GAN for biosignal data augmentation,” *arXiv preprint arXiv:2206.13676*, 2022. [Scopus-indexed via IEEE EMBC proceedings]
- [26] M. W. Lim, S. J. Kim, Y. Park, and N. Kwon, “A deep learning-based time series model with missing value handling techniques to predict various types of liquid cargo traffic,” *Expert Systems with Applications*, vol. 184, p. 115532, 2021, doi: 10.1016/j.eswa.2021.115532.