

PREDIKSI KEBUTUHAN STOK OBAT MENGGUNAKAN METODE HYBRID LONG SHORT-TERM MEMORY (LSTM) DAN CATBOOST

Parwati¹⁾, Muhammad Faisal²⁾, Muhyiddin A.M Hayat³⁾, Nurnawaty⁴⁾, Andi Makbul Syamsuri⁵⁾, M. Agusalim⁶⁾

1. Informatika, Universitas Muhammadiyah Makassar
email: 105841100622@student.unismuh.ac.id
2. Informatika, Universitas Muhammadiyah Makassar
email: muhfaisal@unismuh.ac.id
3. Informatika, Universitas Muhammadiyah Makassar
email: muhyiddin@unismuh.ac.id
4. Teknik Pengairan, Universitas Muhammadiyah Makassar
email: nurnawaty@unismuh.ac.id
5. Teknik Pengairan, Universitas Muhammadiyah Makassar
email: amakbulsyamsuri@unismuh.ac.id
6. Teknik Pengairan, Universitas Muhammadiyah Makassar
email: m.agusalim@unismuh.ac.id

Abstract

Drug inventory planning in primary health facilities requires an accurate forecasting model because fluctuating demand can trigger stockouts or excess inventory. This study develops and evaluates a hybrid Long Short-Term Memory (LSTM) and CatBoost model for predicting the stock requirements of five essential medicines at Puskesmas Pattingalloang. The dataset consists of monthly drug dispensing records from January 2019 to December 2025. LSTM is applied as a temporal feature extractor with a three-month sliding window, while CatBoost functions as the final nonlinear regression estimator. Model performance is assessed using MAE, RMSE, MAPE, and SMAPE, with a single LSTM model used as the baseline comparison. The results show that model suitability depends on the demand pattern of each medicine. The hybrid LSTM-CatBoost model performs better on highly fluctuating medicines, particularly Paracetamol 500 mg with 24.03% SMAPE and Guaifenesin with 37.28% SMAPE. In contrast, the single LSTM model is more efficient for relatively stable demand, especially Blood Supplement Tablets with 14.09% SMAPE. Forecasting for 2026 also provides annual demand estimates that can support data-driven drug requirement planning. These findings indicate that machine learning-based forecasting is useful for pharmaceutical inventory decision support, but model selection must consider the fluctuation characteristics of each drug.

Kata Kunci : CatBoost, LSTM, metode hybrid, ***persediaan*** obat, prediksi, SMAPE, time series

A. PENDAHULUAN

Pengelolaan stok obat merupakan bagian penting dalam mutu pelayanan kesehatan karena ketersediaan obat menentukan kelancaran terapi pasien. Ketidaktepatan perencanaan persediaan dapat memunculkan dua risiko sekaligus, yaitu kelebihan persediaan yang meningkatkan biaya penyimpanan dan risiko kedaluwarsa, serta kekosongan stok yang dapat menghambat pelayanan. Kondisi ini menjadi lebih rumit ketika permintaan obat bergerak dinamis, tidak selalu *linear*, dan dipengaruhi oleh pola penyakit masyarakat dari waktu ke waktu [1], [2].

Pendekatan peramalan konvensional masih sering digunakan dalam perencanaan kebutuhan obat. Akan tetapi, metode statistik klasik cenderung terbatas ketika berhadapan dengan lonjakan permintaan dan pola *non-linear* yang kompleks pada data farmasi [3]. Perkembangan *deep learning* menawarkan alternatif yang lebih adaptif. *Long Short-Term Memory* (LSTM) dikenal mampu mempelajari dependensi temporal dalam data deret waktu, sehingga sesuai untuk prediksi kebutuhan obat yang memiliki keterkaitan antarperiode [4], [5].

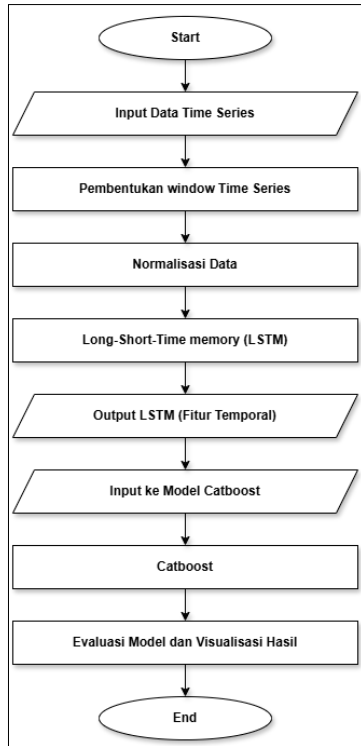
Meski demikian, penggunaan LSTM sebagai model tunggal tidak selalu menjadi pilihan terbaik. Performa LSTM sangat dipengaruhi oleh karakteristik data, jumlah observasi, dan pengaturan parameter. Pada sisi lain, algoritma *boosting* seperti *CatBoost* memiliki kemampuan kuat dalam memodelkan relasi *non-linear* dan mengurangi bias prediksi melalui mekanisme *ordered boosting* [6], [7], [8], [9]. Celah yang kemudian muncul adalah bagaimana menggabungkan kekuatan LSTM dalam mengekstraksi pola temporal dengan *CatBoost* sebagai estimator regresi akhir untuk meningkatkan prediksi kebutuhan stok obat.

Berdasarkan celah tersebut, penelitian ini membangun model *hybrid LSTM-CatBoost* untuk memprediksi kebutuhan stok obat di Puskesmas Pattingalloang. Fokus penelitian diarahkan pada lima jenis obat esensial, yaitu Amoksisillin 500 mg, Guaifenesin, Klorfeniramina Maleat (CTM) 4 mg, Parasetamol 500 mg, dan Tablet Tambah Darah. Tujuan penelitian bukan hanya mencari model dengan nilai galat terendah, tetapi juga menilai kecocokan model terhadap karakteristik permintaan tiap obat.

B. METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimental kuantitatif berbasis data deret waktu. Data yang digunakan berupa data historis pengeluaran obat bulanan pada Puskesmas Pattingalloang selama 84 bulan, mulai Januari 2019 sampai Desember 2025. Variabel utama yang diprediksi adalah jumlah obat keluar, karena variabel tersebut merepresentasikan kebutuhan aktual obat pada periode tertentu.

Tahap awal penelitian meliputi pembersihan data, penyusunan data berdasarkan urutan bulan, dan normalisasi menggunakan *Min-Max Scaling* agar nilai tiap obat berada pada rentang yang sebanding. Setelah itu, data dibentuk menjadi struktur *time series* dengan pendekatan *sliding window*. Berdasarkan analisis ACF dan PACF, pengaruh murni yang paling kuat berada pada lag 1 sampai lag 3, sehingga *window size* ditetapkan sebesar tiga bulan. Artinya, kebutuhan tiga bulan sebelumnya digunakan untuk memprediksi kebutuhan bulan berikutnya [10].



Gambar 1. Alur model hybrid LSTM-CatBoost

Model hybrid dibangun dengan menempatkan LSTM sebagai *feature extractor*. *Hidden state* yang dihasilkan LSTM digunakan sebagai representasi fitur temporal dari pola kebutuhan obat. Fitur tersebut kemudian menjadi input bagi *CatBoost* untuk melakukan regresi akhir. Dengan rancangan ini, LSTM menangkap ketergantungan waktu, sedangkan *CatBoost* memetakan hubungan *non-linear* yang tersisa pada fitur hasil ekstraksi LSTM [11].

Sebagai pembandingan, penelitian ini juga menguji model LSTM tunggal dengan parameter pengujian yang disamakan. Data dibagi menggunakan skema *hold-out split*, yaitu 80% data latih dan 20% data uji. Evaluasi dilakukan menggunakan *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE), dan *Symmetric Mean Absolute Percentage Error* (SMAPE). SMAPE dipakai sebagai metrik penting

karena MAPE dapat menjadi bias ketika nilai aktual mendekati nol [12], [13].

Formula evaluasi yang digunakan Adalah:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

$$SMAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|) / 2}$$

Tahap akhir penelitian melakukan *recursive forecasting* untuk memproyeksikan kebutuhan obat Januari sampai Desember 2026 dengan melatih ulang model menggunakan seluruh data historis 2019-2025.

C. HASIL DAN PEMBAHASAN

Karakteristik Data Kebutuhan Obat

Data penelitian menunjukkan bahwa lima obat memiliki perilaku permintaan yang tidak sama. Parasetamol 500 mg dan Guafenesin memperlihatkan fluktuasi ekstrem dengan lonjakan tajam pada bulan-bulan tertentu. Kondisi ini sesuai dengan karakter medis keduanya yang banyak dipengaruhi keluhan demam, nyeri, batuk, dan peningkatan kasus saluran pernapasan. Sebaliknya, tablet tambah darah cenderung lebih stabil dan teratur karena penggunaannya relatif rutin pada program pencegahan anemia.

Analisis ACF memperlihatkan adanya korelasi berantai pada data kebutuhan obat, sehingga data tidak dapat diperlakukan sebagai rangkaian acak.

Analisis PACF menunjukkan bahwa hubungan langsung yang paling bermakna berada pada lag 1 sampai 3. Temuan ini menjadi dasar penentuan *window size* tiga bulan pada model LSTM. Keputusan ini penting karena *window* yang terlalu pendek dapat kehilangan informasi temporal, sedangkan *window* yang terlalu panjang berpotensi menambah *noise* dan memperberat pelatihan model.

Evaluasi Model Hybrid dan LSTM Tunggal

Hasil pengujian menunjukkan bahwa model *hybrid* LSTM-CatBoost tidak selalu menjadi model paling unggul untuk semua obat. Pada obat dengan fluktuasi ekstrem, model *hybrid* memberikan kesalahan lebih rendah. Namun, pada obat yang memiliki pola relatif stabil, LSTM tunggal justru lebih efisien. Perbandingan berbasis SMAPE disajikan pada Tabel 1.

Tabel 1. Perbandingan SMAPE model *hybrid* dan LSTM Tunggal.

Obat	Hyb.	LSTM	Unggul
Paraset. 500	24,03 %	28,60 %	Hybrid
Guaifenesin	37,28 %	41,79 %	Hybrid
Amoksi. 500	23,82 %	19,66 %	LSTM
CTM 4	34,49 %	31,35 %	LSTM
TTD	16,14 %	14,09 %	LSTM

Tabel 1 memperlihatkan bahwa *hybrid* LSTM-CatBoost unggul pada Parasetamol 500 mg dengan SMAPE 24,03% dan Guaifenesin dengan SMAPE 37,28%. Dua obat ini memiliki pola yang fluktuatif dan cenderung *non-linear*. CatBoost sebagai estimator akhir mampu membantu model memetakan pola

lonjakan yang tidak sepenuhnya tertangkap oleh LSTM tunggal.

Sebaliknya, LSTM tunggal menghasilkan SMAPE lebih rendah pada Amoksisillin 500 mg, CTM 4 mg, dan Tablet Tambah Darah. Pada Tablet Tambah Darah, LSTM tunggal mencapai SMAPE 14,09%, lebih rendah dibanding *hybrid* sebesar 16,14%. Temuan ini mengoreksi asumsi sederhana bahwa model yang lebih kompleks pasti lebih baik. Untuk data yang stabil, model kompleks dapat menambah risiko *overfitting*, sedangkan model tunggal yang lebih sederhana justru cukup untuk menangkap pola temporal.

Nilai MAPE pada beberapa obat juga memperlihatkan kelemahan ketika data aktual sangat kecil. Pada CTM 4 mg, MAPE model *hybrid* mencapai 113,87%, sementara SMAPE berada pada 34,49%. Perbedaan ini menunjukkan bahwa SMAPE lebih adil untuk data kebutuhan obat yang pada periode tertentu dapat mendekati nol. Dengan demikian, interpretasi akurasi tidak boleh hanya bertumpu pada satu metrik, terutama untuk data farmasi yang memiliki karakter pengeluaran tidak merata.

Prediksi Kebutuhan Obat Tahun 2026

Setelah model dievaluasi, penelitian melakukan prediksi kebutuhan obat untuk periode Januari sampai Desember 2026. Hasil prediksi bulanan kemudian diakumulasi menjadi estimasi kebutuhan tahunan. Rekapitulasi total kebutuhan obat pada tahun 2026 disajikan pada Tabel 2.

Tabel 2. Rekapitulasi prediksi total kebutuhan obat tahun 2026

Nama Obat	Total 2026
Amoksisillin 500 mg	27.562
Guaifenesin	27.873

CTM 4 mg	34.206
Parasetamol 500 mg	56.233
Tablet Tambah Darah	57.345

Tabel 2 menunjukkan bahwa kebutuhan tertinggi diproyeksikan terjadi pada Tablet Tambah Darah sebanyak 57.345 tablet dan Parasetamol 500 mg sebanyak 56.233 tablet. Estimasi ini dapat digunakan sebagai *baseline* dalam penyusunan Rencana Kebutuhan Obat (RKO), bukan sebagai angka mutlak yang meniadakan pertimbangan lapangan. Dalam praktik logistik farmasi, hasil prediksi tetap perlu dikombinasikan dengan *safety stock*, catatan wabah, jumlah kunjungan pasien, dan kebijakan program kesehatan.

Secara aplikatif, hasil penelitian menegaskan bahwa sistem prediksi berbasis *machine learning* dapat mendukung pengambilan keputusan persediaan obat secara lebih rasional. Namun, penerapannya tidak boleh bersifat otomatis tanpa interpretasi. Fasilitas kesehatan perlu memilih model berdasarkan karakter data obat: *hybrid LSTM-CatBoost* lebih sesuai untuk obat dengan permintaan tidak stabil, sedangkan LSTM tunggal dapat dipertahankan untuk obat dengan pola reguler.

D. KESIMPULAN DAN SARAN

Penelitian ini berhasil membangun model prediksi kebutuhan stok obat menggunakan pendekatan *hybrid LSTM-CatBoost*. Integrasi dilakukan dengan menggunakan LSTM sebagai pengekstraksi fitur temporal berbasis *sliding window* tiga bulan, kemudian fitur hasil LSTM diproses oleh *CatBoost* sebagai model regresi akhir. Hasil evaluasi menunjukkan bahwa performa model sangat bergantung pada karakteristik permintaan obat.

Model *hybrid LSTM-CatBoost* lebih unggul pada obat yang memiliki fluktuasi ekstrem, yaitu Parasetamol 500 mg dengan SMAPE 24,03% dan Guaifenesin dengan SMAPE 37,28%. Sebaliknya, LSTM tunggal lebih optimal pada obat dengan pola stabil, terutama Tablet Tambah Darah dengan SMAPE 14,09%. Dengan demikian, kebaruan utama penelitian ini bukan hanya pada integrasi dua algoritma, melainkan pada pembuktian bahwa pemilihan model prediksi stok obat harus mempertimbangkan struktur fluktuasi data.

Berdasarkan hasil tersebut, fasilitas kesehatan disarankan menggunakan prediksi sebagai *baseline* penyusunan RKO dan menambahkan *safety stock* sekitar 10-20% untuk mengantisipasi lonjakan kebutuhan yang tidak terduga. Penelitian selanjutnya perlu memperluas pendekatan menjadi *multivariate time series* dengan memasukkan variabel jumlah kunjungan pasien, musim, kejadian wabah, dan kebijakan program kesehatan agar model lebih sensitif terhadap perubahan permintaan.

E. REFERENSI

- [1] K. M. Siregar, Z. Fitri, dan Fajriana, "Prediksi Stok Obat TB Dengan ARIMA dan Analisis Volatilitas Residual di Puskesmas Banda Sakti," *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 726-740, 2025, doi: 10.36341/rabit.v10i2.6398.
- [2] V. A. Lestari, A. Y. Ananta, dan P. Basudewa, "Sistem Informasi Prediksi Persediaan Obat Di Apotek Naylun Farma Menggunakan Holt-Winters," *Jurnal Informatika Polinema*, vol. 9, no. 2, pp. 229-236, 2023, doi: 10.33795/jip.v9i2.1289.
- [3] A. Srinivas, A. Soni, G. Prajapati, dan M. P. S. Manjunath, "Pharmaceutical Sales Data Prediction Using Time

- Series Forecasting,” *International Journal of Intelligent Systems and Applications in Engineering*, vol. 13s, pp. 681-696, 2024.
- [4] M. L. Ashari dan M. Sadikin, “Prediksi Data Transaksi Penjualan Time Series Menggunakan Regresi LSTM,” *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 9, no. 1, p. 1, 2020, doi: 10.23887/janapati.v9i1.19140.
- [5] A. Billah, A. Yolanda, D. Khairani, D. T. Putra, dan H. T. Syif, “Prediksi Jumlah Obat Menggunakan Jaringan Syaraf Tiruan RNN pada Data Penjualan Bulan Juli,” *Jurnal Sains dan Teknologi (JSIT)*, vol. 5, no. 3, pp. 215-221, 2025, doi: 10.47233/jsit.v5i3.3588.
- [6] R. Rathipriya, A. A. Abdul Rahman, S. Dhamodharavadhani, A. Meero, dan G. Yoganandan, “Demand forecasting model for time-series pharmaceutical data using shallow and deep neural network model,” *Neural Computing and Applications*, vol. 35, no. 2, pp. 1945-1957, 2023, doi: 10.1007/s00521-022-07889-9.
- [7] A. M. Abed, “Accelerate demand forecasting by hybridizing CatBoost with the dingo optimization algorithm to support supply chain conceptual framework precisely,” *Frontiers in Sustainability*, vol. 5, 2024, doi: 10.3389/frsus.2024.1388771.
- [8] A. V. Dorogush, V. Ershov, dan A. Gulin, “CatBoost: gradient boosting with categorical features support,” 2018.
- [9] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, dan A. Gulin, “CatBoost: Unbiased boosting with categorical features,” *Advances in Neural Information Processing Systems*, 2020, pp. 6638-6648.
- [10] R. J. Hyndman dan G. Athanasopoulos, *Forecasting: Principles and Practice*. OTexts, 2021.
- [11] H. Hewamalage, C. Bergmeir, dan K. Bandara, “Recurrent Neural Networks for Time Series Forecasting: Current status and future directions,” *International Journal of Forecasting*, vol. 37, no. 1, pp. 388-427, 2021, doi: 10.1016/j.ijforecast.2020.06.008.
- [12] C. J. Willmott dan K. Matsuura, “Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance,” *Climate Research*, vol. 30, no. 1, pp. 79-82, 2020, doi: 10.3354/cr030079.
- [13] S. Makridakis, E. Spiliotis, dan V. Assimakopoulos, “Statistical and Machine Learning forecasting methods: Concerns and ways forward,” *PLoS ONE*, vol. 13, no. 3, e0194889, 2020.
- [14] M. T. Hidayat dan M. Sulistiyono, “Analisis Performa Algoritma XGBoost, GRU, dan Prophet dalam Peramalan Penjualan Obat untuk Optimasi Rantai Pasok Farmasi,” *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 1, pp. 65-73, 2025, doi: 10.52436/1.jpti.562.