

Penerapan LSTM untuk Prediksi Jangka Pendek Penjualan Paket Data (Studi Kasus: Kota Malang)

Muhammad Azkiya Akmal¹, Trimono², Alfian Rizaldy Pratama³

^{1,2,3}*Sains Data, Universitas Pembangunan Nasional "Veteran" Jawa Timur*

¹22083010084@student.upnjatim.ac.id

²trimono.stat@upnjatim.ac.id

³alfan.fasilkom@upnjatim.ac.id

Corresponding author email: 22083010084@student.upnjatim.ac.id

ABSTRAK

Abstrak— Peningkatan aktivitas digital yang pesat menyebabkan volume transaksi penjualan paket data berfluktuasi secara dinamis dalam interval waktu yang pendek. Peramalan data berfrekuensi tinggi (high-frequency data) ini menjadi tantangan tersendiri karena metode statistik konvensional sering kali kurang responsif terhadap pola non-linear dan volatilitas yang tinggi. Penelitian ini bertujuan untuk menerapkan metode Long Short-Term Memory (LSTM) dalam memprediksi penjualan paket data per jam di wilayah Malang, menggunakan data transaksi operasional dari Januari hingga Juni 2024. Kinerja model LSTM dievaluasi dan dibandingkan dengan model statistik SARIMAX menggunakan metrik Mean Absolute Error (MAE) dan Symmetric Mean Absolute Percentage Error (SMAPE). Hasil pengujian menunjukkan bahwa model LSTM mampu menangkap pola temporal kompleks dan fluktuasi tajam dengan lebih baik dibandingkan SARIMAX. Secara kuantitatif, LSTM menghasilkan akurasi yang lebih tinggi dengan nilai MAE sebesar 10.59 dan SMAPE 36.26%, mengungguli SARIMAX yang mencatatkan MAE sebesar 11.32 dan SMAPE 44.29%. Penelitian ini menyimpulkan bahwa pendekatan deep learning lebih efektif dibandingkan metode statistik tradisional untuk mendukung pengambilan keputusan operasional pada data time series non-stasioner di sektor telekomunikasi.

Keywords: Deep Learning, LSTM, Prediksi Penjualan, SARIMAX, Time Series.

I. PENDAHULUAN

Dalam beberapa tahun terakhir, kemajuan teknologi digital telah mendorong lonjakan signifikan dalam aktivitas berbasis internet, terutama di sektor telekomunikasi. Permintaan terhadap layanan data seluler terus meningkat seiring dengan penetrasi smartphone, media sosial, dan layanan streaming yang masif. Misalnya, berita yang ditulis oleh Ericsson menyebutkan bahwa jumlah konsumsi data rata-rata per smartphone di dunia pada 2024 diperkirakan mencapai 21 GB per bulan, hal tersebut merupakan peningkatan empat kali lipat dari tahun 2018 yang diperkirakan karena meningkatnya waktu menonton video, media sosial, dan konten resolusi tinggi [1]. Kondisi ini menyebabkan volume transaksi penjualan paket data meningkat secara drastis dan berlangsung secara dinamis dalam interval waktu yang sangat pendek. Kemampuan memprediksi penjualan secara akurat dalam jangka waktu singkat menjadi penting untuk mendukung keputusan operasional, seperti strategi promosi, serta pengelolaan stok produk digital.

Namun, peramalan penjualan dengan data frekuensi tinggi menghadirkan tantangan metodologis tersendiri. Model statistik konvensional seperti *Autoregressive Integrated*

Moving Average (ARIMA), *Seasonal ARIMA* (SARIMA), maupun SARIMA dengan variabel eksogen (SARIMAX) banyak digunakan dalam penelitian sebelumnya untuk memodelkan data deret waktu musiman. Model-model tersebut efektif untuk data dengan pola linear dan interval harian atau bulanan, tetapi kurang mampu menangkap pola non-linear dan dinamika jangka pendek yang kompleks dalam data transaksi modern. Beberapa literatur menunjukkan bahwa penggunaan data berfrekuensi tinggi (seperti jam atau tiap menit) membuat proses *seasonal adjustment* menjadi sangat menantang [2], [3]. Salah satu keterbatasan utama dari model ARIMA beserta turunannya adalah asumsi stasionaritas dan linearitas hubungan antar variabel. Ketika diterapkan pada data berdimensi tinggi seperti penjualan per jam dengan pengaruh faktor eksternal (hari libur, jam puncak, promosi mendadak), model ini cenderung kehilangan akurasi dan gagal menangkap perubahan perilaku konsumen yang fluktuatif [4].

Penelitian terbaru menunjukkan bahwa pendekatan berbasis *machine learning* dan *deep learning*, khususnya *Recurrent Neural Networks* (RNN) dan turunannya seperti *Long Short-Term Memory* (LSTM), mampu mengatasi keterbatasan model statistik tradisional dengan cara mempelajari pola dependensi jangka panjang serta hubungan non-linear dalam deret waktu [5]. LSTM dirancang untuk menanggulangi masalah *vanishing gradient* pada RNN klasik, sehingga lebih stabil dalam menangani data sekuensial dengan horizon waktu yang panjang. Berbeda dengan SARIMAX yang menekankan pada formulasi model eksplisit berbasis lag dan *differencing*, LSTM secara otomatis mengekstraksi fitur temporal dari data mentah tanpa perlu transformasi kompleks, menjadikannya lebih adaptif untuk konteks data *real-time* dan *high-frequency forecasting* [6].

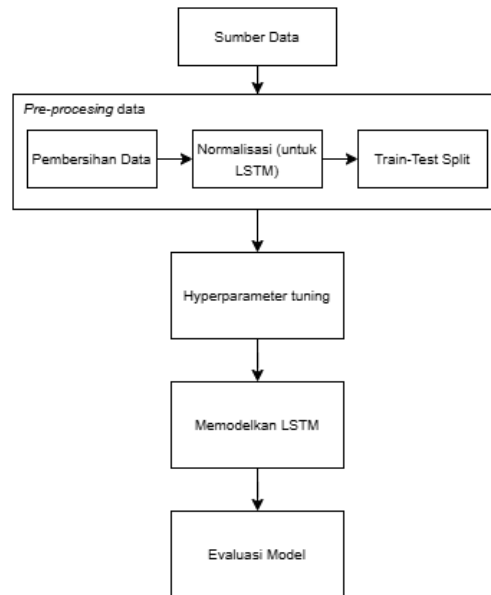
Berbagai penelitian sebelumnya telah membuktikan bahwa *deep learning* mampu menangani data time series yang tidak stasioner, dengan variabilitas dan fluktuasi tinggi [7], [8]. Selain itu, model hibrida yang menggabungkan LSTM dengan teknik klasik seperti ARIMA atau model lain juga menunjukkan peningkatan yang signifikan dalam menangkap pola non-musiman dan anomali [9]. Meskipun demikian, penelitian yang mengimplementasikan LSTM secara spesifik pada domain telekomunikasi Indonesia, khususnya pada konteks penjualan paket data dengan resolusi waktu per jam, masih sangat terbatas.

Keterbatasan penelitian terdahulu yang menggunakan model statistik seperti SARIMAX terletak pada dua hal utama yaitu rendahnya kemampuan generalisasi terhadap data dengan volatilitas tinggi dan noise dan ketergantungan terhadap asumsi eksogen yang sulit dikonfirmasi secara empiris. Dalam praktiknya, fluktuasi penjualan data seluler tidak hanya dipengaruhi oleh faktor kalender seperti hari libur atau akhir pekan, tetapi juga oleh faktor dinamis seperti pola perilaku pengguna, promosi mendadak, hingga pergeseran jam puncak konsumsi. Model statistik konvensional sulit beradaptasi terhadap kondisi semacam ini karena parameter modelnya bersifat tetap setelah proses fitting. Studi yang fokus pada peramalan *high-frequency time series* juga menunjukkan bahwa model konvensional sering memiliki deviasi besar ketika digunakan untuk data frekuensi tinggi [10]. Sebaliknya, LSTM bersifat fleksibel dan mampu belajar secara langsung dari distribusi data historis, menjadikannya lebih cocok untuk data kompleks, bersifat non-stasioner, dan berfrekuensi tinggi.

Penelitian ini bertujuan untuk menerapkan model *Long Short-Term Memory* (LSTM) dalam memprediksi penjualan paket data jangka pendek di kota Malang. Studi ini berfokus pada penerapan metodologi modern berbasis *deep learning* untuk memvalidasi efektivitas LSTM dalam menangani *time series* granular (hourly) dibandingkan pendekatan statistik tradisional. Dengan memanfaatkan data transaksi penjualan aktual dari kanal distribusi resmi, penelitian ini diharapkan dapat memberikan pemahaman empiris mengenai pola temporal dan volatilitas permintaan, sekaligus menjadi dasar pengembangan model hibrida di masa depan yang menggabungkan interpretabilitas model klasik dengan akurasi model *neural network*.

II. METODOLOGI PENELITIAN

Proses penelitian dibagi menjadi beberapa tahap, dimulai dari pengumpulan dan pra-pemrosesan data, pembentukan dataset deret waktu, pembangunan model SARIMAX dan LSTM, hingga evaluasi kinerja dan analisis komparatif terhadap hasil prediksi. Tahapan penelitian ini diuraikan secara sistematis mengikuti alur yang digambarkan pada Gambar 1, yang menjelaskan keseluruhan proses mulai dari persiapan data hingga tahap peramalan dan analisis hasil.



Gambar 1. Alur Kerja Penelitian

1. Data dan Sumber Data

Dataset yang digunakan dalam studi ini diperoleh dari database internal penjualan paket data Telkomsel Regional Jawa Timur, mencakup periode Januari hingga Juni 2024. Dataset tersebut terdiri dari log transaksi per jam, menghasilkan sekitar 4.300 observasi. Setiap baris mewakili total aktivitas penjualan paket data Telkomsel dalam satu jam, memberikan resolusi temporal yang detail dan cocok untuk peramalan jangka pendek menggunakan metode *deep learning*. Dalam penelitian ini, jumlah total transaksi “jumlah_transaksi” dipilih sebagai variabel target yang akan diprediksi, karena secara langsung mencerminkan pola permintaan pelanggan. Fitur tambahan berbasis waktu dan kalender (seperti jam, hari dalam minggu, dan indikator hari libur) dimasukkan untuk membantu model mempelajari variasi periodik dan yang dipicu oleh peristiwa dalam perilaku penjualan.

2. Pra-proses Data

Sebelum pemodelan dilakukan, data harus diluruskan agar bebas dari duplikasi, missing values, outlier ekstrem serta interval waktu yang tidak konsisten. Penanganan *missing values* adalah bagian penting dari pra-pemrosesan *time series* agar model tidak belajar pola artefak karena data kosong atau pengisian yang salah [11]. Karena model LSTM menggunakan optimisasi berbasis *gradient descent* dan jaringan saraf rekuren, maka skala fitur (*feature scaling*) sangat penting agar pelatihan menjadi stabil dan konvergensi cepat. Studi menunjukkan bahwa normalisasi menjadi tahap krusial saat menggunakan jaringan saraf pada data *time series* keuangan, karena rentang variabel yang sangat bervariasi dapat menyebabkan pelatihan yang tidak stabil [12]. Dalam penelitian ini, variabel numerik seperti “jumlah_transaksi” dan fitur tambahan seperti jam

(hour) di-normalisasi menggunakan metode Min-Max Scaling ke rentang [0, 1] dengan persamaan berikut:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

di mana x adalah nilai asli, x_{min} dan x_{max} adalah nilai minimum dan maksimum dari kolom tersebut di data training. Fitur kategorikal seperti *day_of_week*, *is_weekend*, *lbr_nasional*, dan *lbr_minggu* tetap digunakan dalam bentuk biner.

Setelah fitur siap, tahap selanjutnya adalah membagi dataset menjadi data pelatihan (*training*) dan pengujian (*test*) berdasarkan urutan waktu agar tidak terjadi kebocoran data ke depan. Dalam penelitian deret waktu dengan model LSTM maupun SARIMAX, sangat penting menjaga urutan waktu tanpa acak agar skenario peramalan mencerminkan kondisi nyata. Dalam penelitian ini, data dibagi ke dalam proporsi 80% untuk training dan 20% untuk testing, berdasarkan urutan waktu berurutan dari periode awal hingga akhir. Tidak dilakukan pengacakan acak.

3. Hyperparameter Tuning

Setelah data siap digunakan, langkah berikutnya adalah optimasi *hyperparameter* untuk memperoleh konfigurasi model LSTM yang paling tepat. *Hyperparameter* seperti jumlah unit LSTM, nilai *dropout*, *batch size*, dan *learning rate* dievaluasi melalui proses tuning menggunakan pendekatan eksperimental. Tuning dilakukan dengan mencoba berbagai kombinasi parameter untuk melihat konfigurasi yang menghasilkan nilai eror terkecil pada data validasi. Proses ini bertujuan memastikan bahwa model tidak hanya mampu melakukan penyesuaian data latih dengan baik, tetapi juga memiliki kemampuan generalisasi yang baik terhadap data baru.

4. Pemodelan LSTM (Long Short-Term Memory)

Tahap pemodelan dilakukan setelah konfigurasi terbaik diperoleh. Model LSTM digunakan karena kemampuannya untuk mempelajari hubungan jangka panjang dalam deret waktu serta menangani pola non-linear yang sering muncul dalam data penjualan. LSTM bekerja berdasarkan mekanisme *gate* seperti *forget gate*, *input gate*, dan *output gate*. Salah satu komponen penting model ini adalah pembaruan state memori, yang dirumuskan sebagai berikut:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (2)$$

di mana f_t adalah *forget gate*, i_t adalah *input gate*, dan $\tilde{c}_t$ adalah *candidate memory*. Arsitektur model terdiri dari satu lapisan LSTM dengan jumlah unit yang ditentukan dari proses *tuning*, kemudian diikuti oleh lapisan *dropout* untuk mencegah *overfitting*. Keluaran model diproses melalui lapisan *dense* untuk menghasilkan prediksi nilai jumlah transaksi pada jam berikutnya. Proses pelatihan menggunakan fungsi *loss Mean Squared Error* (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

optimizer Adam yang dipilih karena stabil dan efektif dalam mempercepat konvergensi.

5. Evaluasi Model

Tahap terakhir dalam proses penelitian adalah evaluasi performa model LSTM menggunakan dua metrik utama, yaitu *Mean Absolute Error* (MAE) dan *Symmetric Mean Absolute Percentage Error* (SMAPE). MAE digunakan untuk mengukur rata-rata kesalahan absolut antara nilai aktual dan prediksi tanpa mempertimbangkan arah kesalahan. Metrik ini memberikan gambaran seberapa besar deviasi rata-rata prediksi terhadap nilai aktual dalam satuan asli. MAE dihitung menggunakan rumus:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

Selain MAE, penelitian ini juga menggunakan SMAPE untuk mengevaluasi performa model dalam bentuk persentase yang bersifat simetris. SMAPE lebih stabil dibandingkan MAPE pada data yang memiliki nilai aktual mendekati nol, sehingga cocok digunakan pada data penjualan berfrekuensi tinggi yang kadang menampilkan nilai rendah pada jam-jam tertentu. Rumus SMAPE adalah sebagai berikut:

$$SMAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{\left(\frac{|y_i| + |\hat{y}_i|}{2}\right)} \quad (5)$$

Penggunaan dua metrik ini memberikan pemahaman yang lebih komprehensif mengenai performa model, baik dari sisi kesalahan absolut maupun kesalahan relatif. Evaluasi dilakukan menggunakan data testing untuk memastikan bahwa model benar-benar diuji pada data yang tidak pernah dilihat selama proses pelatihan, sehingga hasil evaluasi mencerminkan kemampuan model dalam melakukan peramalan di kondisi nyata.

III. HASIL DAN ANALISIS

Bab ini menyajikan hasil pemodelan LSTM yang digunakan untuk memprediksi jumlah transaksi penjualan paket data per jam. Seluruh proses pelatihan dilakukan menggunakan konfigurasi *hyperparameter* terbaik yang diperoleh pada tahap sebelumnya.

1. Hasil Pelatihan Model LSTM

Pada tahap ini, model LSTM dilatih menggunakan data yang telah dibentuk dalam format sekuens dengan panjang 168 jam. Struktur model terdiri dari satu lapisan LSTM berjumlah 32 unit, diikuti oleh *dropout* dan satu lapisan dense sebagai *output*. Proses pelatihan dilakukan selama 44 *epoch* hingga model mencapai tingkat konvergensi yang stabil. Hasil pelatihan disajikan melalui tabel arsitektur model, bentuk data input, dan kurva loss yang menggambarkan perkembangan error selama pelatihan.

Tabel 1. Bentuk Data Pelatihan dan Pengujian

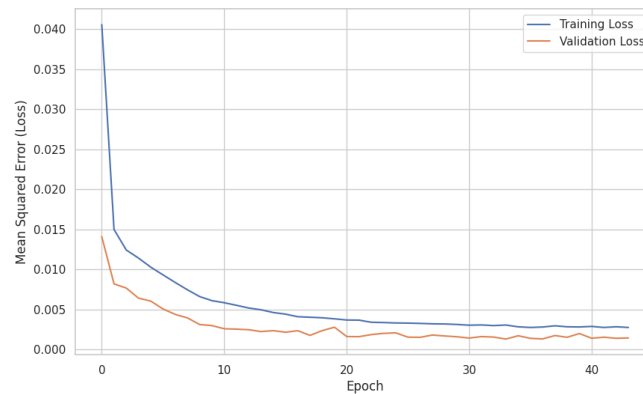
Dataset	Dimensi	Keterangan
X_train	(3480, 168, 6)	3480 sekuens training
y_train	(3480,)	Target training
X_test	(552, 168, 6)	552 sekuens testing
x_test	(552,)	Target testing

Data memiliki format tiga dimensi sesuai kebutuhan LSTM, yaitu *samples* $\times$ *time steps* $\times$ *features*. Panjang sekuens 168 jam memungkinkan model mempelajari pola temporal mingguan. Total fitur per time step sebanyak enam menunjukkan bahwa model memanfaatkan variabel target historis serta lima fitur eksogen lainnya.

Tabel 2. Arsitektur Model LSTM

Layer	Output Shape	Parameter
LSTM (32 units)	(None, 32)	4,992
Dropout (0.2)	(None, 32)	0
Dense (1 unit)	(None, 1)	33
Total Parameter	—	5,025

Arsitektur ini termasuk ringan dengan hanya 5.025 parameter, sehingga proses pelatihan berlangsung cepat dan efisien. *Output* dari LSTM berupa representasi temporal berdimensi 32 yang kemudian diproyeksikan menjadi satu nilai prediksi melalui lapisan dense. Selama pelatihan, nilai *training loss* dan *validation loss* dipantau untuk melihat stabilitas pembelajaran model.

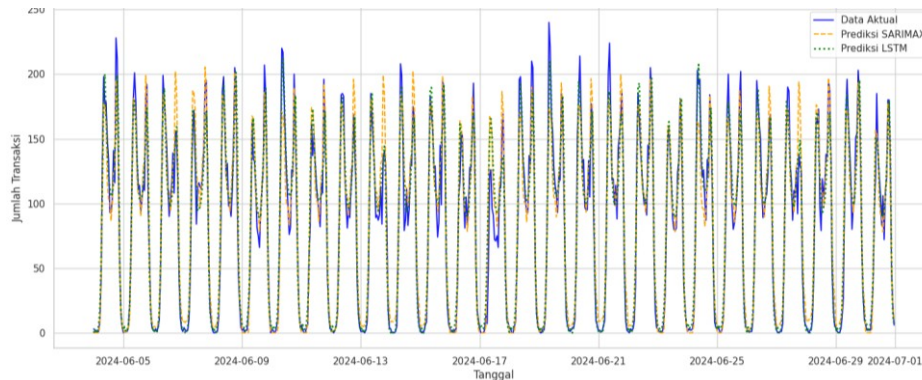


Gambar 2. Kurva Loss Model LSTM

Kurva menunjukkan bahwa *loss* menurun secara konsisten seiring bertambahnya *epoch*. *Validation loss* berada di bawah *training loss* dan stabil setelah sekitar epoch ke-20, mengindikasikan bahwa model tidak mengalami *overfitting*. Nilai *loss* yang semakin kecil menunjukkan bahwa model mampu mempelajari pola deret waktu secara efektif.

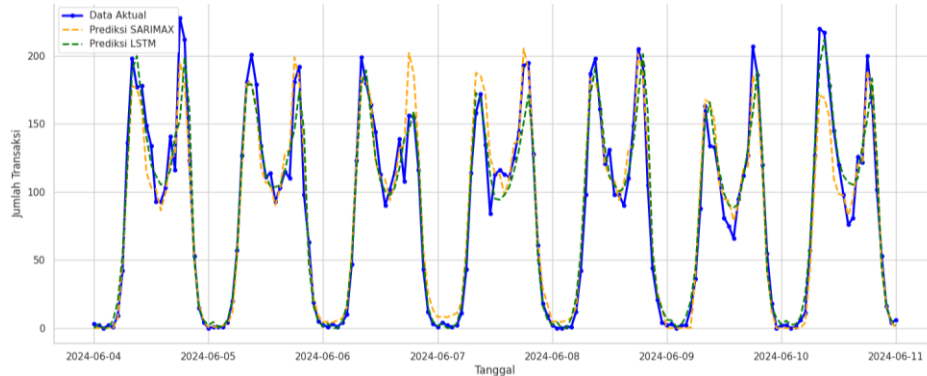
2. Hasil Hyperparameter Tuning

Setelah proses pelatihan selesai, model LSTM dan SARIMAX digunakan untuk melakukan prediksi pada periode pengujian, yaitu bulan Juni 2024. Hasil prediksi kemudian dibandingkan dengan nilai aktual untuk menilai kemampuan model dalam mengikuti pola deret waktu. Visualisasi prediksi serta perhitungan metrik evaluasi digunakan sebagai dasar analisis performa model. Gambar berikut menunjukkan perbandingan antara nilai aktual penjualan paket data dengan hasil prediksi dari dua model, yaitu SARIMAX dan LSTM, pada seluruh periode pengujian.



Gambar 3. Perbandingan Prediksi Model pada Periode Test

Secara visual, kedua model mampu mengikuti pola siklus harian yang sangat jelas pada data aktual. Namun, model LSTM tampak menghasilkan pola prediksi yang lebih mendekati nilai aktual dibandingkan SARIMAX, terutama pada variasi menengah. SARIMAX cenderung menampilkan puncak yang lebih halus dan terkadang sedikit tertinggal pada perubahan nilai yang tajam. Untuk melihat performa model secara lebih detail, dilakukan visualisasi *zoom-in* pada tujuh hari pertama periode *testing*.



Gambar 4. Perbandingan Zoom-in 7 Hari

Pada rentang waktu ini terlihat lebih jelas bahwa LSTM lebih mampu meniru bentuk gelombang harian yang cepat berubah. Puncak transaksi dan saat transaksi mendekati nol direplikasi dengan lebih baik oleh LSTM. Sebaliknya, SARIMAX masih menunjukkan kecenderungan menghasilkan puncak yang sedikit lebih datar dan kurang responsif terhadap perubahan tajam, khususnya di periode menjelang subuh. Namun demikian, keduanya tetap mengikuti pola musiman harian secara konsisten. Untuk mengukur performa prediksi, digunakan dua metrik evaluasi yaitu MAE dan sMAPE. Metrik ini mengukur seberapa dekat hasil prediksi terhadap nilai aktual dari berbagai perspektif error.

Tabel 3. Perbandingan Metrik Evaluasi Model

Metrik	SARIMAX	LSTM
MAE	11.32	10.59
sMAPE	44.29%	36.26%

Hasil evaluasi menunjukkan bahwa model LSTM mengungguli SARIMAX pada seluruh metrik yang digunakan. MAE LSTM lebih rendah, menandakan error absolut rata-rata yang lebih kecil. sMAPE LSTM lebih unggul signifikan, menandakan kemampuan yang lebih baik dalam meniru pola perubahan relatif pada data. Secara keseluruhan, model LSTM lebih efektif dalam menangkap pola musiman harian yang kompleks pada data penjualan berfrekuensi tinggi dibandingkan model SARIMAX.

IV. KESIMPULAN

Penelitian ini berhasil menerapkan metode Long Short-Term Memory (LSTM) untuk memprediksi penjualan paket data jangka pendek dengan resolusi per jam di wilayah Malang. Berdasarkan rangkaian percobaan dan evaluasi yang telah dilakukan, dapat ditarik beberapa kesimpulan utama sebagai berikut:

1. Model LSTM terbukti mampu mempelajari pola temporal yang kompleks pada data penjualan berfrekuensi tinggi (high-frequency data). Dengan arsitektur model yang menggunakan 32 unit LSTM dan dropout sebesar 0.2, model berhasil mencapai konvergensi yang stabil tanpa mengalami overfitting yang signifikan. Hal ini

menunjukkan bahwa pendekatan deep learning adaptif terhadap karakteristik data penjualan paket data yang dinamis.

2. Hasil evaluasi komparatif menunjukkan bahwa model LSTM mengungguli model statistik konvensional SARIMAX dalam hal akurasi prediksi. Berdasarkan metrik evaluasi pada data uji, LSTM menghasilkan nilai kesalahan yang lebih rendah dengan MAE sebesar 10.59 dan SMAPE sebesar 36.26%. Capaian ini lebih baik dibandingkan model SARIMAX yang mencatatkan MAE sebesar 11.32 dan SMAPE sebesar 44.29%. Penurunan nilai SMAPE yang signifikan menandakan bahwa LSTM lebih handal dalam meminimalkan kesalahan relatif terhadap nilai aktual.
3. Secara visual, LSTM menunjukkan responsivitas yang lebih baik dalam mengikuti fluktuasi data yang tajam, termasuk pada saat jam puncak (peak hours) maupun saat transaksi mendekati nol. Sebaliknya, model SARIMAX cenderung menghasilkan pola prediksi yang lebih landai (halus) dan kurang responsif terhadap perubahan mendadak, terutama pada periode transisi waktu yang cepat.
4. Hasil penelitian ini mengonfirmasi bahwa untuk data time series non-stasioner dengan resolusi waktu per jam di sektor telekomunikasi, pendekatan deep learning (LSTM) lebih efektif dibandingkan metode statistik tradisional berbasis lag (SARIMAX). Kemampuan LSTM mengekstraksi fitur temporal secara otomatis menjadikannya solusi yang tepat untuk mendukung pengambilan keputusan operasional yang membutuhkan presisi tinggi dalam jangka pendek.

Berdasarkan hasil penelitian ini, terdapat beberapa rekomendasi untuk pengembangan penelitian selanjutnya guna meningkatkan kualitas prediksi. Pertama, pengembangan pada sisi feature engineering sangat disarankan, yaitu dengan menambahkan variabel eksogen yang lebih kaya seperti jadwal promosi paket data dan informasi event lokal di kota Malang yang berpotensi memicu lonjakan trafik, serta mempertimbangkan penggunaan metode dekomposisi sinyal untuk membantu model memilah komponen tren dan musiman. Kedua, penelitian mendatang diharapkan dapat mengeksplorasi arsitektur hibrida, khususnya model CNN-LSTM, yang memiliki keunggulan dalam menangkap pola spasial-temporal secara simultan, sehingga diharapkan mampu meningkatkan akurasi prediksi model secara signifikan dibandingkan arsitektur tunggal.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Telkomsel Regional Jawa Timur atas izin dan akses data operasional yang diberikan, sehingga penelitian ini dapat terlaksana dengan menggunakan data aktual yang relevan.

Penulis juga menyampaikan apresiasi dan rasa terima kasih kepada Dosen Pembimbing serta jajarannya pengajar Program Studi Sains Data, Universitas Pembangunan Nasional "Veteran" Jawa Timur, atas bimbingan, diskusi yang konstruktif, serta arahan akademik yang sangat berharga selama penyusunan penelitian dan naskah ini.

REFERENSI

- [1] Ericsson, "Smartphone traffic per subscription will continue to grow, driven by increasing video quality and immersive formats.," Oct. 2018. Accessed: Oct. 27, 2025. [Online]. Available: <https://www.ericsson.com/4ac644/assets/local/reports-papers/mobility-report/documents/2018/emr-nov-2018-streaming-video.pdf>
- [2] B. D. K. Reddy, J. S. Naik, S. V. Kumar, S. Kumar, G. Haritha, and M. R. Reddy, "A Methodological Review on Time Series Forecasting by using ARIMA," 2025, pp. 709–719. doi: 10.2991/978-94-6463-662-8_55.
- [3] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, "A Review of ARIMA vs. Machine Learning Approaches for Time Series

- Forecasting in Data Driven Networks,” *Future Internet*, vol. 15, no. 8, p. 255, Jul. 2023, doi: 10.3390/fi15080255.
- [4] Z. Li, J. Han, and Y. Song, “On the forecasting of high-frequency financial time series based on ARIMA model improved by deep learning,” *J Forecast*, vol. 39, no. 7, pp. 1081–1097, Nov. 2020, doi: 10.1002/for.2677.
- [5] G. Lu *et al.*, “A Survey of Deep Learning for Time Series Forecasting: Theories, Datasets, and State-of-the-Art Techniques,” *Computers, Materials & Continua*, vol. 0, no. 0, pp. 1–10, 2025, doi: 10.32604/cmc.2025.068024.
- [6] X. Kong *et al.*, “Deep learning for time series forecasting: a survey,” *International Journal of Machine Learning and Cybernetics*, vol. 16, no. 7–8, pp. 5079–5112, Aug. 2025, doi: 10.1007/s13042-025-02560-w.
- [7] B. Lim and S. Zohren, “Time-series forecasting with deep learning: a survey,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2194, p. 20200209, Apr. 2021, doi: 10.1098/rsta.2020.0209.
- [8] H. Han *et al.*, “Time series forecasting model for non-stationary series pattern extraction using deep learning and GARCH modeling,” *Journal of Cloud Computing*, vol. 13, no. 1, p. 2, Jan. 2024, doi: 10.1186/s13677-023-00576-7.
- [9] I. Aijaz and P. Agarwal, “A Study on Time Series Forecasting using Hybridization of Time Series Models and Neural Networks,” *Recent Advances in Computer Science and Communications*, vol. 13, no. 5, pp. 827–832, Nov. 2020, doi: 10.2174/1573401315666190619112842.
- [10] D. Ollech and D. Bundesbank, “Economic analysis using higher-frequency time series: challenges for seasonal adjustment,” *Empir Econ*, Aug. 2022, doi: 10.1007/s00181-022-02287-5.
- [11] A. Tawakuli, B. Havers, V. Gulisano, D. Kaiser, and T. Engel, “Survey:Time-series data preprocessing: A survey and an empirical analysis,” *Journal of Engineering Research*, vol. 13, no. 2, pp. 674–711, Jun. 2025, doi: 10.1016/j.jer.2024.02.018.
- [12] D. T. Tran, J. Kannianen, M. Gabbouj, and A. Iosifidis, “Bilinear Input Normalization for Neural Networks in Financial Forecasting,” *ArXiv*, vol. abs/2109.00983, 2021, [Online]. Available: <https://api.semanticscholar.org/CorpusID:237385896>