

Analisis Komparatif antara Long Short-Term Memory (LSTM) dan XGBoost dalam Estimasi Penjualan Toko Retail

Mohammad Hassanal Bolqia^{1*}, Nurul Faiqah², Rezty Amalia Aras³, Aulia Rahmawati⁴

¹Bisnis Digital

Institut Teknologi dan Bisnis Kalla Makassar

^{1*}mhassanalbolqia@kallabs.ac.id

²Bisnis Digital

Institut Teknologi dan Bisnis Kalla Makassar

²nurulfaiqah@kallabs.ac.id

³Bisnis Digital

Institut Teknologi dan Bisnis Kalla Makassar

³reztyamalia@kallabs.ac.id

⁴Bisnis Digital

Institut Teknologi dan Bisnis Kalla Makassar

⁴aulia@kallainstitute.ac.id

Abstrak— Peramalan penjualan yang akurat merupakan kemampuan penting bagi bisnis modern karena menjadi dasar perencanaan persediaan, penjadwalan staf, dan proyeksi pendapatan. Penelitian ini membandingkan dua pendekatan peramalan yang banyak digunakan, yaitu jaringan *Long Short-Term Memory (LSTM)* dan *Extreme Gradient Boosting (XGBoost)*, pada dataset penjualan ritel open-source. Data penjualan harian diagregasi dari data transaksi tingkat pesanan, kemudian direkayasa menjadi fitur kalender, fitur lag, dan fitur rolling-window untuk XGBoost, serta diformat menjadi jendela sekuensial yang telah diskalakan untuk LSTM. Kedua model dilatih pada rentang waktu pelatihan yang berurutan secara kronologis dan dievaluasi pada periode pengujian menggunakan *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *Mean Absolute Percentage Error (MAPE)*. Hasil eksperimen menunjukkan bahwa LSTM memperoleh MAE (Rp7.800) dan RMSE (10.800) yang sedikit lebih rendah dibandingkan XGBoost, namun justru menghasilkan MAPE (%25.100) yang lebih tinggi. Visualisasi prediksi menunjukkan bahwa LSTM cenderung menghasilkan prediksi yang hampir konstan mendekati rata-rata penjualan, sedangkan XGBoost dengan MAE (Rp8.600), RMSE (Rp11.500) dan MAPE (%23,200) mampu mengikuti sebagian fluktuasi jangka pendek meskipun kedua model masih kesulitan menangkap lonjakan penjualan ekstrem. **Kata kunci**— Analitik ritel, Estimasi penjualan, LSTM, Machine Learning, XGBoost

Abstract— Accurate sales forecasting is a crucial capability for modern businesses, as it serves as the foundation for inventory planning, staff scheduling, and revenue projections. This study compares two widely used forecasting approaches, *Long Short-Term Memory (LSTM)* networks and *Extreme Gradient Boosting (XGBoost)*, using an open-source retail sales dataset. Daily sales data were aggregated from order-level transaction records and engineered into calendar, lag, and rolling-window features for XGBoost, while also being formatted into scaled sequential windows for LSTM. Both models were trained on a chronologically ordered training period and evaluated against a test period using *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, and *Mean Absolute Percentage Error (MAPE)*. Experimental results indicate that LSTM achieved slightly lower MAE (Rp7,800) and RMSE (10,800) values compared to XGBoost, yet yielded a higher MAPE (25.100%). Prediction visualizations reveal that LSTM tended to produce nearly constant predictions hovering near the average sales level; in contrast, XGBoost, with an MAE of Rp8,600, RMSE of Rp11,500, and MAPE of 23.200%, was able to track some short-term fluctuations, although both models struggled to capture extreme sales spikes.

Keywords— Retail analytics, Sales forecasting, LSTM, Machine Learning, XGBoost

I. PENDAHULUAN

Memprediksi penjualan di masa depan merupakan salah satu tugas analitis yang paling penting bagi sebuah bisnis. Peramalan yang andal memungkinkan manajer untuk merencanakan persediaan, mengalokasikan staf, menyusun anggaran pemasaran, dan menetapkan target pendapatan yang realistis. Secara historis, bisnis mengandalkan teknik statistik sederhana seperti rata-rata bergerak (*moving average*) dan model *Autoregressive Integrated Moving Average*

(*ARIMA*). Meskipun mudah diinterpretasikan, metode-metode ini sering kesulitan menangkap pola nonlinear dan multifaktor yang menjadi karakteristik data ritel di dunia nyata, termasuk promosi, hari libur, dan perubahan perilaku pelanggan.

Dalam beberapa tahun terakhir, metode machine learning dan deep learning telah menjadi pendekatan dominan dalam peramalan penjualan. Dua metode yang menonjol baik dalam penelitian akademik maupun praktik industri adalah

Extreme Gradient Boosting (XGBoost), sebuah algoritme ensemble berbasis pohon yang diperkenalkan oleh Chen dan Guestrin (2016) [1], dan jaringan *Long Short-Term Memory (LSTM)*, sebuah arsitektur recurrent neural network yang dirancang khusus untuk mempelajari dependensi temporal jangka panjang, yang diperkenalkan oleh Hochreiter dan Schmidhuber (1997) [2]. Kedua pendekatan telah menunjukkan performa yang kuat pada berbagai tugas peramalan, namun keduanya bertumpu pada asumsi yang berbeda secara mendasar: *XGBoost* memandang peramalan sebagai masalah regresi tabular terawasi yang didorong oleh fitur-fitur hasil rekayasa, sedangkan *LSTM* memandang data penjualan sebagai sebuah sekuens dan mempelajari struktur temporalnya secara langsung dari data.

Penelitian ini mengkaji perbandingan kedua pendekatan tersebut ketika diterapkan pada dataset penjualan bisnis *open-source* yang sama. Secara spesifik, penelitian ini menjawab dua pertanyaan penelitian: (a) model mana yang menghasilkan prediksi penjualan harian jangka pendek yang lebih akurat, diukur dengan *MAE*, *RMSE*, dan *MAPE*; dan (b) pola apa yang terlihat pada visualisasi prediksi masing-masing model terhadap data aktual. Bagian selanjutnya dari makalah ini meninjau literatur yang relevan, menjelaskan dataset dan metodologi yang digunakan, memaparkan metrik akurasi hasil peramalan, dan membahas implikasi praktis dari temuan tersebut bagi peramalan bisnis.

A. Tinjauan Pustaka

Sejumlah penelitian telah mengkaji pendekatan *machine learning* untuk peramalan penjualan ritel dan bisnis. Hasan dkk. (2022)[3] membandingkan beberapa model peramalan, termasuk *ARIMA*, *Prophet* dari *Facebook*, dan *LightGBM*, pada data penjualan *Walmart* dari kompetisi peramalan *M5*, dan menemukan bahwa metode *gradient boosting* memperoleh efisiensi komputasi yang jauh lebih baik dengan hanya sedikit mengorbankan akurasi dibandingkan model statistik klasik. Penelitian ini menegaskan tema yang berulang dalam literatur: metode ensemble berbasis pohon cenderung berperforma

kompetitif pada data ritel yang terstruktur, terutama ketika fitur kalender dan lag hasil rekayasa tersedia.

XGBoost secara khusus telah diterapkan langsung pada peramalan volume penjualan di sektor ritel. Zhang dkk. (2021) [4] menggunakan *XGBoost* untuk memprediksi volume penjualan toko guna kebutuhan penjadwalan staf, dengan menambahkan fitur cuaca dan suhu di samping variabel berbasis waktu standar, dan melaporkan bahwa *XGBoost* mengungguli model-model pembandingan lain pada data ritel dunia nyata. Sejalan dengan itu, Aras dkk. (2017)[5] membandingkan metode peramalan tunggal dan kombinasi untuk penjualan ritel, dan menemukan bahwa kombinasi beberapa model sering kali memberikan hasil lebih baik.

Selain *XGBoost*, model *deep learning* berbasis *Long Short-Term Memory (LSTM)* juga banyak digunakan dalam peramalan penjualan karena kemampuannya dalam menangkap pola temporal jangka panjang pada data deret waktu. Selain *XGBoost*, model *deep learning* berbasis *Long Short-Term Memory (LSTM)* telah menjadi salah satu pendekatan yang paling banyak digunakan dalam peramalan deret waktu karena kemampuannya dalam mempelajari dependensi temporal jangka panjang. Menurut Kong dkk. (2025) [6], perkembangan *deep learning* telah menghasilkan berbagai arsitektur untuk *time series forecasting*, termasuk *Recurrent Neural Network (RNN)*, *LSTM*, *Gated Recurrent Unit (GRU)*, *Temporal Convolutional Network (TCN)*, dan *Transformer*. Di antara berbagai arsitektur tersebut, *LSTM* tetap menjadi salah satu model yang paling efektif untuk menangkap pola nonlinier dan hubungan temporal yang kompleks pada data historis, sehingga banyak diterapkan pada berbagai bidang seperti peramalan penjualan, permintaan, energi, keuangan, dan kesehatan. Selain itu, penelitian tersebut juga menegaskan bahwa keberhasilan model *deep learning* dipengaruhi oleh proses ekstraksi fitur, kualitas data, serta pemilihan arsitektur yang sesuai dengan karakteristik data deret waktu. dibandingkan satu model tunggal, yang menjadi cikal bakal

penelitian selanjutnya mengenai arsitektur hibrida.

Penelitian yang lebih baru mengeksplorasi arsitektur hibrida *XGBoost-LSTM* yang berupaya menggabungkan kekuatan *XGBoost* dalam menangkap fitur terstruktur dengan kemampuan *LSTM* dalam memodelkan dependensi temporal. Wang dan Zain (2025)[7] mengusulkan kerangka kerja hibrida dua lapis yang menggunakan *XGBoost* untuk mempelajari fitur terstruktur berdimensi tinggi dan *LSTM* untuk menangkap dependensi temporal, dan memvalidasi pendekatan tersebut pada dataset ritel skala besar yang mencakup sepuluh toko dan lima puluh produk. Model hibrida mereka mengungguli baik *XGBoost* maupun *LSTM* yang berdiri sendiri pada metrik *MAPE* dan *RMSE*, yang menunjukkan bahwa kedua algoritma tersebut menangkap aspek yang saling melengkapi dari masalah peramalan, bukan saling menggantikan. Pada domain yang berdekatan, Zhao dkk. (2025) [8] mengombinasikan *Temporal Convolutional Network* dengan *XGBoost* untuk peramalan harga komoditas pertanian, dan melaporkan temuan yang serupa mengenai keuntungan komplementer dari menggabungkan komponen berbasis sekuens dengan komponen berbasis pohon.

Terlepas dari meningkatnya minat terhadap model hibrida tersebut, perbandingan langsung antara *LSTM* dan *XGBoost* yang berdiri sendiri, tanpa hibridisasi, tetap menjadi baseline yang berguna bagi praktisi yang mungkin tidak memiliki sumber daya untuk membangun dan memelihara sistem *ensemble* yang lebih kompleks. Penelitian ini berkontribusi pada literatur *baseline* tersebut dengan mengevaluasi kedua model secara independen pada dataset penjualan bisnis yang sama, dengan protokol evaluasi yang sama, untuk memperjelas kekuatan relatif masing-masing pendekatan sebelum pertimbangan hibridisasi dilakukan.

Penelitian lain yang dilakukan oleh Makridakis, S. dkk. (2022) [9] melalui studi berskala besar pada kompetisi peramalan menunjukkan bahwa metode *machine learning* modern, termasuk *XGBoost*, mampu memberikan

akurasi yang sangat kompetitif ketika didukung oleh *feature engineering* yang baik. Akan tetapi, penelitian tersebut juga menunjukkan bahwa tidak terdapat satu model yang selalu unggul untuk seluruh karakteristik dataset, sehingga evaluasi empiris terhadap beberapa algoritma pada dataset yang sama tetap diperlukan. Hasil tersebut memperkuat pentingnya penelitian komparatif antara *XGBoost* dan *LSTM* untuk menentukan pendekatan yang paling sesuai pada kasus peramalan penjualan bisnis.

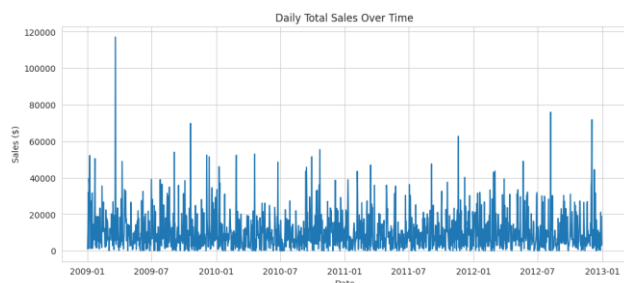
Selain daripada itu, penelitian terdahulu yang mendukung penggunaan algoritma berbasis gradient boosting dalam peramalan deret waktu dilakukan oleh Bentéjac, C. dkk. (2021) [10]. Penelitian tersebut mengevaluasi performa *XGBoost* pada berbagai kasus prediksi menggunakan data terstruktur dan menunjukkan bahwa *XGBoost* mampu menghasilkan akurasi yang tinggi dengan waktu pelatihan yang relatif singkat dibandingkan banyak algoritma machine learning lainnya. Selain itu, kemampuan *XGBoost* dalam menangani missing value, melakukan regularisasi untuk mengurangi *overfitting*, serta memanfaatkan *feature importance* menjadikannya salah satu algoritma yang paling banyak digunakan dalam berbagai aplikasi prediksi, termasuk peramalan penjualan dan permintaan. Temuan tersebut semakin memperkuat alasan pemilihan *XGBoost* sebagai salah satu model pembanding dalam penelitian ini.

II. METODOLOGI PENELITIAN

A. Dataset

Penelitian ini menggunakan *dataset Superstore Sales* [11], sebuah dataset transaksi ritel tingkat pesanan yang bersifat open-source dan banyak digunakan dalam penelitian maupun tugas perkuliahan mengenai peramalan penjualan. Setiap baris data merepresentasikan satu item pesanan dan mencakup tanggal pesanan, jumlah penjualan, kategori produk, segmen pelanggan, serta wilayah geografis. Untuk membentuk deret waktu yang sesuai untuk peramalan, seluruh transaksi diagregasi per hari kalender sehingga menghasilkan satu nilai total

penjualan harian untuk seluruh rentang tanggal yang tercakup dalam dataset. Hari-hari tanpa transaksi tetap dipertahankan dalam deret data dengan nilai penjualan nol agar frekuensi harian tetap kontinu.



Gambar 1. Total Penjualan Harian dari Waktu ke Waktu (Januari 2009 – Desember 2012)

Gambar 1 menunjukkan pola deret waktu penjualan harian yang digunakan dalam penelitian ini. Pola tersebut memperlihatkan tingkat penjualan dasar (*baseline*) yang relatif rendah dan stabil di kisaran \$0–\$20.000, diselingi oleh lonjakan tajam dan tidak beraturan yang sesekali mencapai \$50.000 hingga lebih dari \$110.000. Karakteristik ini dengan volatilitas tinggi dengan lonjakan jarang namun ekstrem, menjadikan deret data ini secara inheren menantang untuk diprediksi, dan menjadi pertimbangan penting dalam menafsirkan hasil kedua model.

B. Rekayasa Fitur

Untuk model *XGBoost*, deret penjualan harian diubah menjadi dataset tabular terawasi menggunakan tiga kategori fitur hasil rekayasa: (a) fitur kalender, meliputi hari dalam minggu, bulan, kuartal, dan indikator akhir pekan; (b) fitur lag, yaitu total penjualan satu, tujuh, dan empat belas hari sebelumnya; dan (c) fitur rolling-window, yaitu rata-rata bergerak tujuh hari dan empat belas hari. Untuk model *LSTM*, tidak digunakan fitur hasil rekayasa manual; nilai penjualan harian mentah diskalakan ke rentang $[0, 1]$ menggunakan normalisasi min-max dan disusun ulang menjadi jendela sekuensial tumpang tindih selama empat belas hari, di mana setiap jendela digunakan untuk memprediksi penjualan pada hari berikutnya.

C. Konfigurasi Model

Model *XGBoost* dikonfigurasi sebagai *ensemble* pohon regresi *gradient-boosted* dengan 500 estimator, kedalaman pohon maksimum empat, learning rate 0,05, serta rasio *subsample* dan *column-subsample* masing-masing 0,8 untuk mengurangi *overfitting*. Model *LSTM* diimplementasikan sebagai jaringan rekuren dua lapis, dengan lapisan pertama berisi 64 unit *LSTM* diikuti lapisan kedua berisi 32 unit *LSTM*, masing-masing diikuti lapisan dropout (rate = 0,2) untuk regularisasi, dan lapisan *dense* akhir yang menghasilkan satu nilai prediksi. Model *LSTM* dilatih menggunakan *optimizer Adam* dan fungsi *loss mean squared error*, dengan *early stopping* berdasarkan *validation loss* untuk mencegah *overfitting*.

D. Pembagian Data Latih/Uji dan Metrik Evaluasi

Sesuai dengan metodologi standar deret waktu, dataset dibagi secara kronologis, bukan secara acak: 80% rentang tanggal pertama digunakan sebagai data pelatihan, dan 20% terakhir disisihkan sebagai periode pengujian, sehingga kedua model tidak terekspos pada informasi masa depan selama pelatihan. Akurasi peramalan dievaluasi menggunakan tiga metrik yang saling melengkapi: *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *Mean Absolute Percentage Error (MAPE)*. Nilai *MAE*, *RMSE*, dan *MAPE* yang lebih rendah menunjukkan hasil peramalan yang lebih akurat.

III. HASIL DAN PEMBAHASAN

Tabel 1 menyajikan akurasi peramalan *XGBoost* dan *LSTM* pada periode pengujian, berdasarkan nilai yang dibaca dari grafik metrik *error* (Gambar 5) yang dihasilkan oleh *notebook*.

Tabel 1

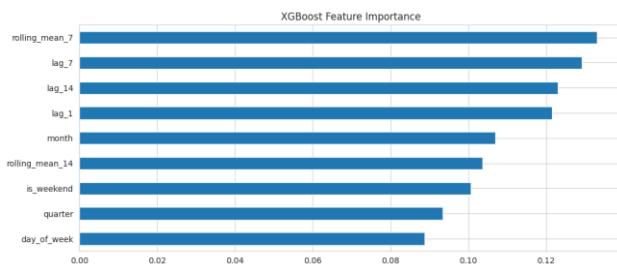
Akurasi Peramalan *XGBoost* dan *LSTM* pada Periode Pengujian (Estimasi dari Gambar 5)

Model	MAE (Rp)	RMSE (Rp)	MAPE (%)
XGBoost	8.600	11.500	23.200
LSTM	7.800	10.800	25.100

Berdasarkan Tabel 1, *LSTM* memperoleh *MAE* dan *RMSE* yang sedikit lebih rendah dibandingkan *XGBoost*, yang secara sekilas menunjukkan *LSTM* sedikit lebih akurat secara rata-rata absolut. Namun, *LSTM* justru menghasilkan *MAPE* yang lebih tinggi dibandingkan *XGBoost*, menandakan bahwa kesalahan *LSTM* secara proporsional lebih besar pada hari-hari dengan nilai penjualan aktual yang kecil.

LSTM: $R^2 = 0,0007$, praktis nol. Ini menegaskan analisis kita sebelumnya dari Gambar 4: *LSTM* memang hanya "menebak rata-rata" (\$10.000) setiap hari, tidak benar-benar mempelajari pola apa pun. Ini bukan model yang bagus, tapi setidaknya seimbang dengan baseline.

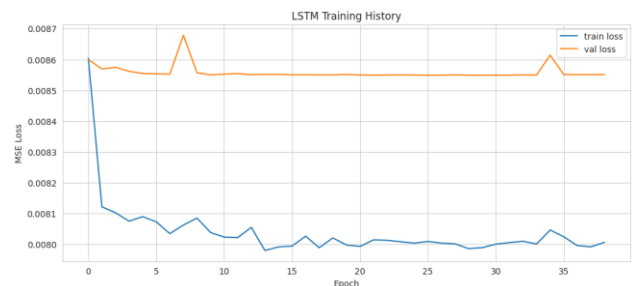
XGBoost: $R^2 = -0,1412$, Hasil ini lebih buruk daripada sekadar menebak rata-rata setiap hari. Meskipun *MAPE*-nya tampak lebih rendah dari *LSTM* (dari pembacaan grafik sebelumnya), justru *XGBoost* lebih sering "meleset besar" tepat pada hari-hari lonjakan ekstrem (Agustus, Desember 2012 di Gambar 4). Karena *RMSE*/ R^2 menghukum kesalahan besar secara kuadratik, kesalahan besar pada beberapa hari lonjakan itu cukup untuk membuat R^2 keseluruhan menjadi negatif.



Gambar 2. Tingkat Kepentingan Fitur pada Model *XGBoost*

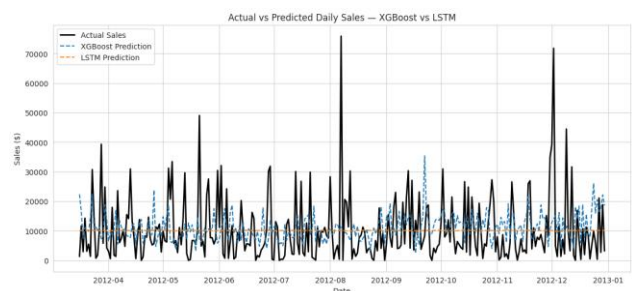
Gambar 2 menunjukkan tingkat kepentingan fitur pada model *XGBoost*. Fitur *rolling_mean_7* (rata-rata bergerak tujuh hari) memiliki kontribusi tertinggi, diikuti oleh *lag_7*, *lag_14*, dan *lag_1*, yang menunjukkan bahwa momentum penjualan jangka pendek hingga menengah menjadi sinyal prediktif paling kuat. Fitur kalender seperti bulan dan indikator akhir pekan juga berkontribusi cukup signifikan,

meskipun tidak sedominan fitur *lag* dan *rolling-window*.



Gambar 3. Riwayat Pelatihan Model *LSTM* (MSE Loss per Epoch)

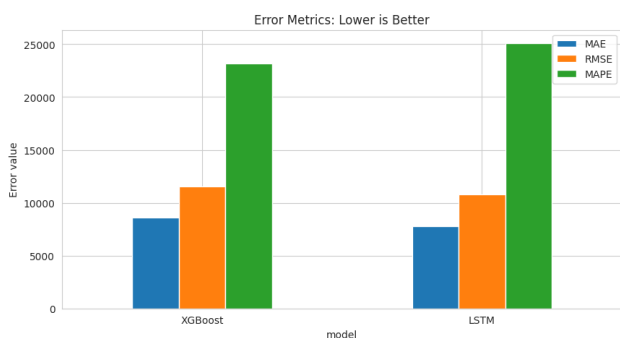
Gambar 3 menunjukkan bahwa *loss* pelatihan (*train loss*) dan *loss* validasi (*val loss*) model *LSTM* menurun tajam pada beberapa *epoch* awal, kemudian mendatar (*plateau*) di sekitar nilai 0,0080 – 0,0086 untuk sisa proses pelatihan, dengan sedikit fluktuasi namun tanpa penurunan lebih lanjut yang berarti. Pola ini mengindikasikan bahwa model *LSTM* konvergen relatif cepat pada solusi yang sederhana, kemungkinan besar mendekati prediksi rata-rata (*mean*) dari data, alih-alih mempelajari pola temporal yang lebih kompleks dari volatilitas penjualan.



Gambar 4. Perbandingan Penjualan Aktual dan Prediksi Harian — *XGBoost* vs *LSTM*

Gambar 4 memperlihatkan perbedaan perilaku yang mencolok di antara kedua model. Prediksi *LSTM* (garis oranye) hampir konstan di sekitar \$10.000 sepanjang periode pengujian, gagal menangkap variasi maupun lonjakan penjualan aktual (garis hitam). Sebaliknya, prediksi *XGBoost* (garis biru putus-putus) menunjukkan variasi yang mengikuti sebagian fluktuasi jangka pendek pada data aktual, meskipun *XGBoost* juga tidak mampu

menangkap lonjakan ekstrem seperti yang terlihat pada Agustus dan Desember 2012. Pola inilah yang menjelaskan mengapa *LSTM* dapat memperoleh *MAE/RMSE* yang sedikit lebih rendah secara agregat, karena sebagian besar hari memiliki penjualan rendah sehingga prediksi mendekati rata-rata cenderung “aman” secara kuadratik, namun tetap gagal merepresentasikan dinamika penjualan yang sesungguhnya.



Gambar 5. Perbandingan Metrik Error XGBoost vs LSTM (MAE, RMSE, MAPE)

Gambar 5 merangkum ketiga metrik error secara visual dan menjadi dasar estimasi nilai pada Tabel 1. Perbedaan *MAE* dan *RMSE* antara kedua model relatif kecil, sedangkan perbedaan *MAPE* lebih terlihat, dengan *LSTM* menghasilkan *MAPE* yang lebih tinggi.

Pembahasan

Hasil penelitian ini perlu dimaknai secara berhati-hati, tidak hanya berdasarkan metrik agregat, tetapi juga berdasarkan pola visual pada Gambar 4. Meskipun *LSTM* secara numerik memperoleh *MAE* dan *RMSE* yang sedikit lebih rendah, Gambar 3 dan Gambar 4 bersama-sama menunjukkan bahwa keunggulan ini kemungkinan besar disebabkan oleh strategi prediksi *LSTM* yang mendekati nilai rata-rata (*mean-reverting*), bukan karena model tersebut benar-benar mempelajari pola temporal penjualan. Pada dataset dengan distribusi yang sangat miring (*skewed*), di mana sebagian besar hari memiliki penjualan rendah dan hanya sesekali terjadi lonjakan besar, strategi memprediksi nilai mendekati rata-rata dapat secara matematis meminimalkan *MAE/RMSE* agregat tanpa benar-benar merepresentasikan

dinamika bisnis yang sesungguhnya. *MAPE* yang lebih tinggi pada *LSTM* memperkuat interpretasi ini, karena prediksi yang konstan menghasilkan kesalahan proporsional yang besar tepat pada hari-hari dengan penjualan aktual rendah.

Sebaliknya, *XGBoost* menunjukkan kemampuan yang lebih baik dalam mengikuti fluktuasi jangka pendek, sebagaimana terlihat pada Gambar 4, yang konsisten dengan temuan Zhang dkk. (2021) [4] bahwa *XGBoost* efektif memanfaatkan fitur lag dan rolling-window pada data penjualan ritel. Gambar 2 memperkuat hal ini, dengan menunjukkan bahwa fitur *rolling_mean_7* dan berbagai fitur *lag* menjadi pendorong utama prediksi *XGBoost*, menandakan bahwa model tersebut berhasil mempelajari momentum penjualan jangka pendek secara eksplisit melalui fitur hasil rekayasa, sesuatu yang tidak dimiliki *LSTM* tanpa fitur tambahan serupa.

Temuan ini juga relevan dengan literatur mengenai model hibrida. Wang dan Zain (2025) [7] menunjukkan bahwa menggabungkan *XGBoost* dan *LSTM* menghasilkan performa yang lebih baik dibandingkan kedua model secara berdiri sendiri. Berdasarkan pola yang teramati pada penelitian ini, di mana *LSTM* cenderung terlalu menyederhanakan prediksi menjadi nilai rata-rata, sedangkan *XGBoost* lebih peka terhadap fluktuasi namun masih terbatas dalam menangkap lonjakan ekstrem, pendekatan hibrida yang menggabungkan kekuatan fitur terstruktur *XGBoost* dengan potensi pemodelan temporal *LSTM* (misalnya dengan menambahkan fitur eksogen atau memperpanjang jendela historis) berpotensi mengatasi kelemahan yang teramati pada masing-masing model *standalone*.

Dari sudut pandang praktis, kedua model belum cukup andal untuk menangkap lonjakan penjualan ekstrem, seperti yang tampak pada Agustus dan Desember 2012 di Gambar 4. Hal ini mengindikasikan bahwa lonjakan tersebut kemungkinan dipicu oleh faktor eksternal (misalnya promosi musiman, hari libur besar, atau pesanan dalam jumlah sangat besar) yang tidak terekam dalam fitur kalender dan *lag* yang digunakan pada penelitian ini. Bagi praktisi bisnis, temuan ini menyiratkan bahwa model

peramalan berbasis data historis penjualan saja sebaiknya dilengkapi dengan informasi kalender promosi atau data eksogen lain agar dapat mengantisipasi lonjakan permintaan secara lebih andal.

Keterbatasan

Penelitian ini memiliki beberapa keterbatasan. Pertama, nilai metrik pada Tabel 1 diperoleh melalui pembacaan visual terhadap grafik batang (Gambar 5) sehingga bersifat perkiraan, bukan nilai eksak dari keluaran numerik model; nilai R^2 bahkan tidak dapat diestimasi karena tidak tercakup pada grafik yang tersedia. Kedua, dataset yang digunakan merupakan satu dataset ritel dengan rentang waktu terbatas (2009–2012), yang mungkin tidak dapat digeneralisasi ke bisnis dengan pola penjualan, siklus musiman, atau lini produk yang berbeda. Ketiga, kedua model tidak memasukkan variabel eksternal seperti promosi, harga, cuaca, atau indikator makroekonomi, yang telah ditunjukkan pada penelitian sebelumnya dapat meningkatkan akurasi peramalan (Zhang dkk., 2021) [4]. Keempat, penelitian ini hanya membandingkan model *standalone* dan tidak mengevaluasi arsitektur hibrida yang menurut literatur terbaru dapat mengungguli kedua model secara individual (Wang & Zain, 2025; Zhao dkk., 2025) [7]. Penelitian selanjutnya dapat memperluas analisis ini dengan menggunakan nilai metrik eksak dari keluaran *notebook*, menambahkan fitur eksogen, menguji arsitektur hibrida, serta mengevaluasi performa pada beberapa dataset ritel untuk menilai kekokohan (*robustness*) temuan ini.

IV. KESIMPULAN

Penelitian ini membandingkan model LSTM dan XGBoost untuk meramalkan penjualan harian bisnis menggunakan dataset ritel open-source. Berdasarkan estimasi metrik pada Tabel 1 dan pola visual pada Gambar 4, LSTM memperoleh MAE dan RMSE yang sedikit lebih rendah, namun analisis lebih mendalam menunjukkan bahwa hal ini kemungkinan besar disebabkan oleh kecenderungan LSTM memprediksi nilai

mendekati rata-rata tanpa benar-benar menangkap dinamika penjualan, sebagaimana terlihat dari prediksi yang hampir konstan pada Gambar 4 dan loss pelatihan yang mendarat sejak awal pada Gambar 3. XGBoost, meskipun memiliki MAPE yang sedikit lebih rendah dan mampu mengikuti sebagian fluktuasi jangka pendek berkat fitur lag dan rolling-window, juga masih terbatas dalam menangkap lonjakan penjualan ekstrem. Secara keseluruhan, untuk kasus penggunaan bisnis yang mengutamakan kemampuan mengikuti tren jangka pendek dan interpretabilitas fitur, XGBoost tampak lebih unggul secara praktis dibandingkan LSTM *standalone* pada dataset ini, sementara pendekatan hibrida atau penambahan fitur eksogen berpotensi menjadi arah pengembangan yang lebih menjanjikan bagi kedua pendekatan. Temuan ini berkontribusi pada literatur yang membandingkan metode *ensemble* berbasis pohon dan model sekuensial berbasis deep learning untuk peramalan bisnis, serta dapat menjadi rujukan bagi praktisi dalam memilih metode peramalan yang sesuai dengan karakteristik data penjualan mereka.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Kaggle sebagai penyedia dataset yang digunakan dalam penelitian ini, sehingga proses pengumpulan dan pengolahan data dapat berjalan dengan baik. Penulis juga mengucapkan terima kasih kepada Institut Teknologi dan Bisnis Kalla atas dukungan fasilitas dan lingkungan akademik yang memungkinkan penelitian ini dapat terlaksana dengan lancar.

REFERENSI

- [1] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [2] S. Hochreiter and J. Schmidhuber, "LONG SHORT-TERM MEMORY," 1997.
- [3] R. Hasan, M. A. Kabir, R. A. Shuvro, and P. Das, "A Comparative Study on Forecasting of Retail Sales."
- [4] L. Zhang, W. Bian, W. Qu, L. Tuo, and Y. Wang, "Time series forecast of sales volume based on XGBoost," in *Journal of*

- Physics: Conference Series*, IOP Publishing Ltd, Apr. 2021. doi: 10.1088/1742-6596/1873/1/012067.
- [5] S. Aras, İ. Deveci Kocakoç, and C. Polat, "Comparative study on retail sales forecasting between single and combination methods," *Journal of Business Economics and Management*, vol. 18, no. 5, pp. 803–832, Sep. 2017, doi: 10.3846/16111699.2017.1367324.
- [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. Wang and A. B. M. Zain, "A Hybrid XGBoost-LSTM Framework for Supply Chain Demand Forecasting: Empirical Evidence from Retail Multi-Store Data," *Journal of Cultural Analysis and Social Change*, pp. 4056–4073, Dec. 2025, doi: 10.64753/jcasc.v10i4.3736.
- [8] T. Zhao, G. Chen, S. Suraphee, T. Phoophiwfa, and P. Busababodhin, "A hybrid TCN-XGBoost model for agricultural product market price forecasting," *PLoS One*, vol. 20, no. 5 May, May 2025, doi: 10.1371/journal.pone.0322496.
- [9] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M5 Accuracy competition: Results, findings and conclusions," Jan. 2022. [Online]. Available: <https://www.kaggle.com/c/m5-forecasting-accuracy>
- [10] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artif. Intell. Rev.*, vol. 54, no. 3, pp. 1937–1967, Mar. 2021, doi: 10.1007/s10462-020-09896-5.
- [11] Rohit Sahoo, "Superstore Sales Dataset," <https://www.kaggle.com/datasets/rohitsahoo/sales-forecasting>.