

Bi-LSTM with Explainable AI for Session Duration-Based Customer Lifetime Value Proxy on Multi-Category E-Commerce Platforms

Yulian Dwi Nartriani ^{1*}, De Rosal Ignatius Moses Setiadi ^{2**}

* Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

** Kelompok Keilmuan Komputasi Kuantum dan Informatika Material, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia

yulinartriani137@gmail.com¹, moses@dsn.dinus.ac.id²

Article Info

Article history:

Received 2025-10-25

Revised 2025-11-26

Accepted 2025-12-10

Keyword:

*Bi-LSTM,
Multi-Category E-Commerce,
Explainable Artificial Intelligence,
Customer Lifetime Value
Prediction,
Customer Session Duration.*

ABSTRACT

The rapid growth of multi-category e-commerce platforms has increased the importance of behavioral data for predicting Customer Lifetime Value (CLV). However, monetary-based CLV estimation is often infeasible due to incomplete or unavailable transaction records. This study adopts session duration as a short-term behavioral proxy for CLV and proposes a Bidirectional Long Short-Term Memory (Bi-LSTM) model enhanced with a Temporal Attention mechanism to improve predictive accuracy. The publicly available REES46 dataset, consisting of 1,6 million events and 276.000 unique sessions, is used with preprocessing steps including label encoding, temporal feature construction, and outlier-aware sampling to address the highly right-skewed distribution of session durations. Four baseline models Decision Tree, Random Forest, Extreme Gradient Boosting (XGBoost), and conventional Long Short-Term Memory (LSTM) are implemented for comparative evaluation. The baseline LSTM achieves MAE = 0,0080 and RMSE = 0,0322. The proposed Bi-LSTM v3 model, equipped with Temporal Attention and structured sampling, demonstrates substantial performance improvement, achieving MAE = 0,0043 (\approx 368 seconds) and RMSE = 0,0172 (\approx 1466 seconds), representing an accuracy gain of approximately 45–50% over the baseline. Explainability analysis using SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) confirms that the time_diff feature is the dominant contributor at both global and local levels, aligning with the behavior of the attention mechanism. Additionally, the integration of Explainable Artificial Intelligence (XAI) provides transparent insights into model decision patterns. These findings show that combining Bi-LSTM, Temporal Attention, and XAI yields an accurate and interpretable framework for session duration prediction, supporting the use of session duration as a feasible CLV proxy in multi-category e-commerce environments.



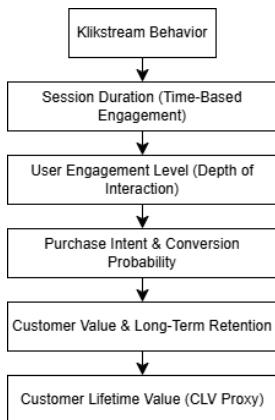
This is an open access article under the [CC-BY-SA](#) license.

I. PENDAHULUAN

Perkembangan pesat *e-commerce* multi-kategori telah menjadikan data perilaku pelanggan sebagai aset strategis, yang kini menjadi sumber utama untuk memahami nilai dan loyalitas pelanggan. Dalam konteks ini, Prediksi *Customer Lifetime Value* (CLV) semakin penting untuk mendukung strategi akuisisi dan retensi pelanggan terutama melalui integrasi model prediktif berbasis *Artificial Intelligence* (AI) [1], [2]. Namun, keterbatasan data finansial kerap

menghambat penghitungan CLV secara langsung. Untuk mengatasi hal ini, sejumlah penelitian mulai mengadopsi pendekatan proxy non-moneter, seperti frekuensi kunjungan, intensitas klik, dan durasi sesi pelanggan (session duration), sebagai alternatif estimasi nilai pelanggan [3], [4]. Durasi sesi, secara khusus, telah terbukti menjadi indikator yang kuat terhadap tingkat keterlibatan (*engagement*) dan loyalitas, terutama dalam ekosistem digital berbasis pengalaman, seperti *e-commerce*, media streaming, dan aplikasi layanan

edukasi [5]. Studi terbaru menunjukkan bahwa durasi sesi dan ritme interaksi (*time-based events*) memiliki korelasi positif dengan *purchase likelihood* dan perilaku pembelian, menjadikannya proxy yang valid untuk CLV ketika data transaksi tidak tersedia [6]. Temuan ini diperkuat oleh penelitian [7], yang mengidentifikasi *session duration* dan *time since last event* sebagai dua prediktor terpenting dalam model prediksi pembelian berbasis klikstream.



Gambar 1 Kerangka Konseptual Penggunaan Session Duration Sebagai Proxy CLV

Diagram pada Gambar 1. menggambarkan kerangka konseptual penggunaan session duration sebagai proxy CLV, yang menghubungkan dinamika temporal interaksi pengguna dengan estimasi nilai pelanggan jangka panjang dalam konteks perilaku multi-kategori.

Analisis perilaku pelanggan melalui data klikstream menuntut metode yang mampu menangkap pola sekuensial dan dependensi temporal yang kompleks. Sebagian besar studi sebelumnya masih mengandalkan model tabular statis yang terbatas dalam merepresentasikan dinamika perilaku pengguna. Misalnya, [8] menunjukkan pendekatan *classification–optimization* untuk pengalokasian anggaran retensi pelanggan berbasis probabilitas *churn*, namun model tersebut mengabaikan hubungan sekuensial antar aktivitas pengguna. Sementara itu, [9] menunjukkan potensi data klikstream dalam mengidentifikasi pelanggan bernilai tinggi melalui algoritma *Customer Merit (CM) index*, yang mengevaluasi aktivitas, efisiensi pencarian, serta ritme interaksi pengguna Walaupun efektif dalam optimasi bisnis, pendekatan-pendekatan ini belum cukup dalam menangkap hubungan temporal antar interaksi pelanggan secara mendalam.

Penelitian berbasis model sekuensial telah menunjukkan potensi dalam memahami pola perilaku pelanggan secara temporal. Adamu et al. [3] menerapkan model *Decision Tree*, *Long Short-Term Memory (LSTM)*, dan gabungannya dengan interpretasi berbasis *SHAP* untuk menjelaskan pengaruh fitur temporal seperti *begin_second* (waktu mulai sesi), pendekatan tersebut masih berfokus pada fitur temporal global dan belum menangkap *intra-session dynamics* yang lebih granular. Padahal, ritme interaksi antar peristiwa

(*time_diff*) dapat menjadi representasi paling langsung dari tingkat keterlibatan pengguna dalam suatu sesi.

Sejalan dengan hal ini, [5] menunjukkan bahwa pemodelan urutan (*sequence-to-sequence learning*) secara signifikan meningkatkan akurasi prediksi CLV dibandingkan model berbasis fitur statis, karena mampu menangkap hubungan temporal antarinteraksi pelanggan. Disisi lain, studi oleh [10] membuktikan efektivitas *Bidirectional Long Short-Term Memory (Bi-LSTM)* dalam mendeteksi anomali temporal secara real-time pada jaringan pengawasan, di mana mekanisme dua arah pada sistem deteksi anomaly *real-time*. Kedua studi tersebut memperkuat bahwa pendekatan Bi-LSTM unggul dalam memahami pola temporal yang kompleks dan relevan untuk prediksi perilaku dinamis seperti durasi sesi pelanggan.

Meskipun demikian, masih terdapat dua celah penelitian yang belum banyak dikaji. Pertama, sebagian besar penelitian CLV berbasis perilaku belum memanfaatkan mekanisme *temporal attention layer* untuk menyoroti bagian urutan yang paling berpengaruh terhadap prediksi durasi sesi. Kedua, isu *interpretability* atau transparansi model pembelajaran mendalam masih jarang dieksplorasi dalam konteks *e-commerce*. Padahal, model yang akurat tetapi tidak dapat dijelaskan (*black-box model*) berisiko mengurangi kepercayaan pengambil keputusan bisnis terhadap hasil prediksi [11],[12].

Menanggapi keterbatasan tersebut, penelitian ini mengusulkan pengembangan model *Bi-LSTM* yang dikombinasikan dengan *temporal attention* memungkinkan model untuk membaca urutan data dari dua arah, masa lalu dan masa depan, sehingga mampu menangkap konteks temporal secara lebih komprehensif dan mempertahankan ketergantungan jangka panjang dalam sekuens perilaku pengguna [13]. Mekanisme *attention* selanjutnya memberikan bobot adaptif terhadap bagian urutan yang paling relevan, sehingga mempermudah interpretasi model terhadap pola interaksi pengguna [14]. Kombinasi *bidirectional modeling* dan *temporal attention* ini diharapkan mampu merepresentasikan *intra-session dynamics* secara lebih granular, termasuk ritme serta intensitas interaksi antar peristiwa yang merefleksikan tingkat keterlibatan pelanggan dalam platform e-commerce multi-kategori.

Selain itu, strategi *outlier-aware sampling* diterapkan untuk menyeimbangkan distribusi durasi sesi yang umumnya *right-skewed* dan meningkatkan stabilitas generalisasi model [15]. Lebih lanjut penelitian ini juga menekankan transparansi model dengan mengimplementasikan *Explainable Artificial Intelligence (XAI)* menggunakan *SHapley Additive exPlanations (SHAP)* dan *Local Interpretable Model-Agnostic Explanations (LIME)* untuk memberikan interpretasi global dan lokal terhadap hasil prediksi[16], [17], [18].

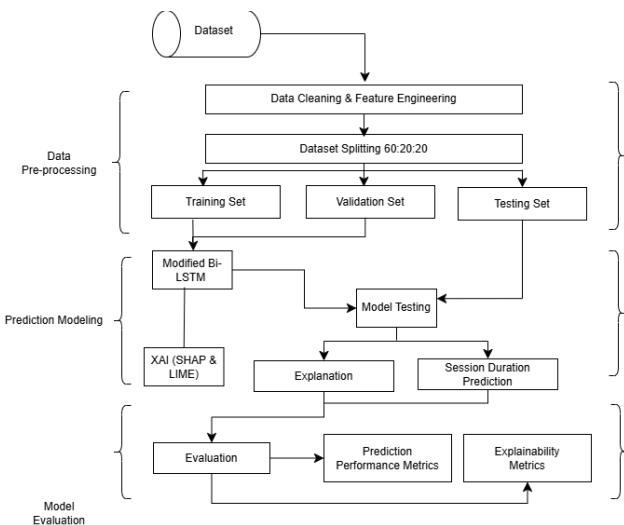
Dengan demikian, tujuan penelitian ini adalah untuk: (1) mengembangkan model Bi-LSTM dengan mekanisme perhatian temporal yang mampu merepresentasikan dinamika intra-sesi secara lebih akurat, dan (2) menerapkan XAI guna

memberikan interpretasi global dan lokal terhadap hasil prediksi CLV berbasis durasi sesi.

Penelitian ini berkontribusi dalam dua dimensi utama: peningkatan presisi prediksi perilaku pelanggan melalui model *deep learning* berbasis durasi sesi serta peningkatan interpretabilitas model prediktif di konteks *e-commerce* multi-kategori.

II. METODE

Penelitian ini mengikuti kerangka *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang mencakup empat tahap utama: *data understanding*, *data preparation*, *modeling*, dan *evaluation* sebagaimana ditunjukkan pada Gambar 2. Fokus utama penelitian berada pada tahapan modeling, dengan penekanan pada pemodelan sekuensial berbasis temporal untuk memprediksi durasi sesi pelanggan sebagai proxy CLV.



Gambar 2 Kerangka Penelitian Prediksi Durasi Sesi Pelanggan Menggunakan Modified Bi-LSTM dan XAI

A. Dataset dan Karakteristik Data

Dataset yang digunakan adalah *e-Commerce Behavior Data* (EBD) dari REES46 (2020) yang tersedia secara publik di *Kaggle*. Data tersebut dianggap representatif untuk prediksi durasi sesi sebagai proxy CLV, serta telah digunakan dalam penelitian sebelumnya oleh [3], sehingga memungkinkan validasi lintas studi. Dataset ini dipilih karena sumber datanya merefleksikan perilaku pengunjung di aplikasi mobile dan situs web. Mengacu pada [19], log sesi web dan aplikasi dianggap sebagai dataset terlengkap yang tersedia untuk keperluan prediksi CLV.

Untuk memastikan eksperimen dapat direproduksi serta mengatasi keterbatasan memori, penelitian ini menggunakan 100.000 sesi unik yang diambil melalui random session sampling, dan seluruh event dalam setiap sesi dipertahankan agar struktur urutan perilaku tidak terdistorsi. Karena proses sampling bersifat acak, nilai agregat seperti mean, median, dan maksimum durasi sesi dapat mengalami variasi antar-run;

oleh karena itu, file sampel final disimpan sebagai canonical dataset.

a. Distribusi Event-Type

Terdapat tiga jenis interaksi utama, yaitu *view*, *cart*, dan *purchase*. Distribusi frekuensi pada 100.000 sesi unik ditunjukkan pada Tabel 1. Distribusi ini mencerminkan pola klikstream umum dengan dominasi aktivitas *view*.

TABEL I
DISTRIBUSI EVENT-TYPE

Event-Type	Jumlah	Persentase
View	517.921	88,9%
Cart	53.770	9,2%
purchase	10.708	1,8%

b. Distribusi Kategori Produk

Dataset memuat 136 kategori produk. Top-10 kategori disajikan pada Tabel 2, yang menegaskan heterogenitas produk pada platform multi-kategori.

TABEL II
DISTRIBUSI KATEGORI PRODUK

Kategori Produk	Jumlah Event
construction.tools.light	136.114
electronics.audio.headphone	74.591
appliances.personal.massager	21.818
appliances.kitchen.refrigerators	20.679
sport.bicycle	18.571
apparel.shoes	14.685
apparel.shoes.slipons	13.530
furniture.kitchen.table	11.536
electronics.clocks	11.160
appliances.environment.vacuum	10.628

c. Statistik Durasi Sesi

Durasi sesi dihitung dari akumulasi *time_diff* antar-event dalam satu *user_session*. Statistik ringkas ditunjukkan pada Tabel 3 (mean, median, min, max). Distribusi durasi bersifat right-skewed sehingga membutuhkan penanganan outlier.

TABEL III
STATISTIK DURASI SESI

Statistik	Nilai (detik)
Rata - rata	412,3
Median	185,0
Maksimum	2574,1
Minimum	0

d. Panjang Urutan (Jumlah Event per Sesi)

Karakteristik interaksi pengguna dihitung berdasarkan total event per *user_id*. Jumlah Event ditunjukkan pada Tabel 4. berdasarkan distribusi long-tail tersebut, MAXLEN untuk model sekuens ditetapkan sebesar 50 timestep, yang mencakup persentase besar sesi tanpa kehilangan informasi temporal utama.

Secara keseluruhan, beberapa atribut dalam dataset menunjukkan tingkat sparsitas yang wajar, di mana Tingkat missing value pada *category_code* sekitar 9% dan brand sekitar 5%. Atribut *price* mengandung nilai ekstrem pada persentil 99,9%, sehingga memerlukan trimming dan imputasi. Nilai *time_diff* menunjukkan variasi besar (median 16 detik; beberapa >300 detik). Variasi ini menegaskan pentingnya mekanisme *temporal attention* untuk menyoroti event paling relevan dalam satu sesi.

TABEL IV
JUMLAH EVENT PER SESI

Statistik	Jumlah Event
Rata-rata	11,97
Median	4
Maksimum	1.810

Dengan demikian, karakteristik dataset ini memberikan landasan yang kuat bagi pemodelan durasi sesi berbasis pendekatan sekuensial dan mendukung relevansi penggunaan Bi-LSTM dengan mekanisme *temporal attention*.

B. Segmentasi Pengguna dan Kategori Produk

Dataset penelitian mencakup berbagai kategori produk dan menunjukkan variasi perilaku pengguna yang cukup signifikan. Untuk memeriksa heterogenitas tersebut, analisis segmentasi diterapkan guna mengevaluasi perbedaan pola interaksi dan distribusi durasi sesi pada masing-masing kategori serta kelompok pengguna.

Segmentasi kategori dilakukan menggunakan atribut *category_code* untuk melihat perbedaan pola interaksi pada tiap kelompok produk. Ringkasan hasil ditampilkan pada Tabel 5.

TABEL V
RINGKASAN KARAKTERISTIK INTERAKSI BERDASARKAN KATEGORI PRODUK

Kategori Produk	Jumlah Interaksi	Kisaran Harga	Karakteristik Interaksi
<i>electronics.smartphone</i>	Tinggi	Sedang-Tinggi	Eksplorasi panjang, banyak perpindahan antar produk
<i>electronics.audio.headphone</i>	Sedang	Sedang	Interaksi stabil, sering perbandingan produk
<i>appliances.kitchen</i>	Rendah-Sedang	Sedang	Sesi lebih singkat, keputusan cepat
<i>furniture.living_room</i>	Rendah	Tinggi	Interaksi terbatas namun bernilai tinggi

Berdasarkan Tabel 5, dapat dianalisis bahwa kategori elektronik cenderung menghasilkan sesi lebih panjang dibandingkan kategori seperti *kitchen appliances* atau furniture, mengindikasikan bahwa perilaku pengguna sangat bergantung pada jenis produk yang diakses.

Segmentasi pengguna diturunkan dari jumlah sesi per *user_id* dan intensitas interaksi pada setiap sesi. Empat kelompok utama diperoleh sebagaimana diringkas pada Tabel 6.

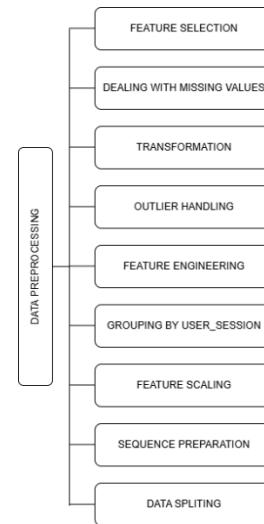
TABEL VI
SEGMENTASI PENGGUNA BERDASARKAN AKTIVITAS

Segmen Pengguna	Definisi	Karakteristik Utama
New User	1 Sesi	Durasi pendek, eksplorasi awal
Occasional User	2-3 Sesi	Aktivitas tidak konsisten
Frequent User	>3 Sesi	Pola browsing stabil dan berulang
High-Engagement User	>10 Event/Sesi	Sesi panjang, interaksi intensif

Hasil segmentasi menunjukkan bahwa tingkat keterlibatan pengguna merupakan faktor yang berpengaruh terhadap panjang durasi sesi. Pengguna dengan interaksi intensif cenderung menghasilkan durasi sesi yang lebih panjang dibandingkan kelompok pengguna lainnya.

Secara keseluruhan, analisis segmentasi ini memberikan gambaran bahwa dataset memiliki variasi perilaku yang cukup beragam, baik dari sisi kategori produk maupun kelompok pengguna. Pola temporal yang muncul pada berbagai segmen menunjukkan bahwa karakteristik *time_diff* tetap menjadi indikator utama yang mencerminkan dinamika durasi sesi. Selain itu, temuan ini membuka peluang penelitian lanjutan melalui pengembangan model prediksi yang lebih spesifik berdasarkan kategori atau segmen pengguna guna meningkatkan akurasi estimasi CLV.

C. Data Preprocessing dan Sampling



Gambar 3 Aktivitas Data Preprocessing

Gambar 3 menunjukkan rangkaian aktivitas preprocessing yang dilakukan pada dataset sebelum digunakan pada proses pemodelan. Pada penelitian ini *feature selection* dilakukan

secara manual (*conceptual-based feature determination*) dengan fokus pada fitur yang relevan terhadap perilaku pelanggan dan dinamika temporal, yaitu *event_time*, *event_type*, *category_code*, *brand*, *user_session*, dan *price*.

Tahapan *cleaning* dilakukan dengan menghapus instance yang mengandung data duplikat dan nilai kosong pada atribut utama (*event_time*, *user_session*, dan *price*) serta mengurutkan event berdasarkan waktu dalam setiap sesi, pada atribut utama. Atribut *event_type*, *category_code*, dan *brand* diubah menjadi representasi numerik menggunakan label encoding.

Sebagai bagian penting dari proses persiapan data, penelitian ini menerapkan skema *outlier-aware sampling* menggunakan pendekatan yang terstruktur. Pertama, outlier harga (*price*) ditangani melalui *quantile-based trimming* dengan menghapus nilai-nilai ekstrem pada persentil 99,9%. Kedua, outlier durasi sesi ditangani dengan *rule-based trimming* dengan menghapus sesi berdurasi lebih dari 24 jam (86.400 detik), berdasarkan evaluasi distribusi durasi yang menunjukkan pola *highly right-skewed*. Ketiga, setelah proses trimming dilakukan, diterapkan *random session sampling* terhadap 100.000 sesi unik untuk menjaga keragaman perilaku pengguna sekaligus memastikan efisiensi komputasi. Langkah-langkah ini memastikan bahwa data yang digunakan model bebas dari nilai ekstrem yang tidak representatif, serta memenuhi prinsip *reproducibility* seperti yang dianjurkan dalam penelitian berbasis klikstream.

Tahapan *feature engineering* menghasilkan delapan fitur utama di tingkat sesi yang mewakili perilaku dan dimensi temporal pengguna yang diuraikan dalam Tabel 4. Fitur *time_diff* dan *time_diff_ratio* diperkenalkan sebagai kontribusi baru penelitian ini untuk merepresentasikan ritme interaksi antar-event. Pendekatan ini memberikan kemampuan tambahan bagi model untuk memahami *temporal flow* perilaku pelanggan, dibandingkan pendekatan konvensional yang hanya mengandalkan fitur waktu global.

TABEL VII
SPESIFIKASI FITUR

No	Fitur	Deskripsi
1	<i>event_type_enc</i>	Jenis aktivitas pengguna (view/cart/purchase)
2	<i>category_code_enc</i>	Kategori produk yang dilihat/dibeli
3	<i>brand enc</i>	Preferensi merek pengguna
4	<i>price</i>	Harga produk
5	<i>time diff</i>	Selisih waktu antar event
6	<i>time_diff_ratio</i>	Proporsi tiap event terhadap durasi session
7	<i>pos in session</i>	Urutan event dalam session
8	<i>daypart (onehot)</i>	Variabel temporal kategori waktu

D. Model Sekuens

Data urutan (*sequence*) dari setiap *user_session* diproses ke dalam bentuk *padded sequence* dengan panjang maksimum (MAXLEN) sebesar 50 timestep. Nilai ini ditetapkan berdasarkan analisis distribusi panjang sesi, di mana 90% sesi memiliki jumlah event ≤ 50 . Penetapan MAXLEN=50

menjaga keseimbangan antara cakupan informasi temporal yang relevan dan efisiensi komputasi, sejalan dengan praktik umum dalam sequence modeling berbasis LSTM. Padding dilakukan di bagian akhir (*post-padding*) untuk mempertahankan urutan alami event.

E. Perhitungan Durasi Sesi

Durasi sesi dihitung sebagai akumulasi selisih waktu antar-event (*time_diff*) dalam setiap *user_session*, bukan menggunakan timestamp awal dan akhir secara langsung. Pendekatan berbasis agregasi ini menegaskan fokus penelitian pada *intra-session dynamics*, yaitu ritme interaksi pengguna dalam satu sesi. Cara ini menghindari bias temporal akibat perbedaan zona waktu atau gangguan sistem logging, sekaligus menjawab kritik terhadap pendekatan konvensional yang hanya menggunakan selisih global antara *start time* dan *end time* seperti pada penelitian Adamu et al [3].

Sebagai catatan metodologis, penggunaan session duration sebagai proxy CLV telah dijustifikasi pada Bab 1 melalui dukungan literatur yang menunjukkan hubungan positif antara durasi sesi, engagement, dan kecenderungan purchase, sehingga pendekatan ini tetap valid meskipun dataset tidak menyediakan data transaksi.

F. Model dan Pelatihan

Baseline model berupa LSTM dan *Decision Tree* direplikasi sepenuhnya dari [3], termasuk konfigurasi *hidden units*, *panjang sequence*, dan *fungsi loss*. Pemilihan baseline ini konsisten dengan penelitian terdahulu mengenai prediksi perilaku pelanggan berbasis klikstream. Selain baseline tersebut, penelitian ini juga memasukkan dua model machine learning klasik yang secara luas digunakan dalam regresi tabular, yaitu *Random Forest* dan *XGBoost*. Penambahan kedua model ini memastikan ketersediaan baseline non-sekuensial yang kuat sehingga evaluasi performa tidak hanya membandingkan arsitektur deep learning, tetapi juga pendekatan prediktif konvensional. Dengan demikian, penelitian ini mencakup empat baseline yang mewakili model sekuensial dan non-sekuensial.

Model utama yang diusulkan dalam penelitian ini adalah *Modified Bi-LSTM*, yang memperluas arsitektur *Bidirectional LSTM* dengan *Temporal Attention Layer*. Komponen ini dirancang untuk memperkuat kemampuan model dalam menyoroti event-event penting dalam urutan waktu, sehingga relevan untuk menangkap dinamika perilaku pelanggan yang kompleks [20]. Bi-LSTM dipilih dibandingkan GRU, Transformer Encoder, atau Temporal Convolutional Network (TCN) karena menawarkan stabilitas dan efisiensi yang lebih baik pada urutan pendek (≤ 50 event), sesuai karakteristik dataset penelitian ini.

Pengembangan model dilakukan secara bertahap dalam tiga versi iteratif, yaitu v1, v2, dan v3. Setiap versi dikembangkan secara modular untuk mengevaluasi kontribusi setiap komponen baru terhadap peningkatan performa prediksi, sehingga berfungsi sebagai *internal ablation study*.

Modified Bi-LSTM v1 (Stacked + Temporal Fusion) merupakan versi awal dengan dua lapisan Bi-LSTM berukuran besar untuk menangkap dependensi temporal dua arah secara lebih dalam.

Modified Bi-LSTM v2 (Huber Loss + Feature Enhancement) menambahkan loss function yang lebih tahan terhadap outlier serta memperkaya input dengan fitur statistik seperti harga rata-rata, maksimum, dan minimum.

Modified Bi-LSTM v3 (Temporal Attention + Outlier-Aware Sampling) versi akhir, dilengkapi lapisan Temporal Attention untuk pembobotan adaptif antar event serta strategi sampling yang mempertimbangkan distribusi sesi ekstrem.

Dari ketiga versi tersebut, v3 dipilih sebagai model final karena menunjukkan performa prediktif terbaik. Strategi pengembangan bertingkat seperti ini selaras dengan pendekatan eksperimental pada arsitektur adaptif seperti yang diterapkan oleh [20], yang menekankan pentingnya integrasi bertahap untuk memaksimalkan relevansi fitur dan akurasi model.

Kontribusi utama arsitektur yang diusulkan terletak pada integrasi *Temporal Attention* dengan Bi-LSTM dua-lapisan yang dirancang secara khusus untuk mempelajari dinamika *time_diff* sebagai ritme perilaku pengguna. Selain itu, strategi *outlier-aware* sampling diterapkan untuk menangani distribusi durasi sesi yang sangat ekstrem, sebuah kondisi yang unik pada dataset klikstream multi-kategori dan belum dijelaskan dalam penelitian sebelumnya.

Konfigurasi hyperparameter dan gambar arsitektur model yang dicantumkan dalam Tabel 8, Tabel 9 dan Gambar 4 difokuskan pada arsitektur Modified Bi-LSTM v3, sebagai model utama yang digunakan dalam eksperimen. Parameter untuk v1 dan v2 tidak ditampilkan secara rinci karena sebagian besar komponennya identik dengan v3 (jumlah fitur input, panjang sequence, optimizer, dan batch size). Perbedaan antar versi hanya pada struktur lapisan dan konfigurasi fungsi loss, yang telah dijelaskan secara naratif pada uraian model. Pendekatan ini lazim digunakan pada eksperimen bertingkat untuk menyoroti kontribusi komponen baru secara sistematis.

TABEL VIII
HYPERPARAMETER DAN DETAIL ARSITEKTUR MODEL BI-LSTM V3

Komponen	Deskripsi	Nilai
Input Shape	Dimensi input sekuens (timestep × fitur)	50 × 5
Jumlah unit LSTM (layer 1 &2)	Jumlah neuron pada tiap arah dalam lapisan Bi-LSTM	192 dan 128
Return Sequences	Konfigurasi keluaran LSTM untuk menjaga urutan antar-layer	Ya (layer 1), Ya (layer 2)
Attention Units	Dimensi vektor pada mekanisme temporal attention	64

Pooling Layer	Jenis pooling yang digunakan setelah Bi-LSTM	Global-Average-Pooling1D
Optimizer	Algoritma optimasi untuk mempercepat konvergensi	Adam
Learning Rate	Laju pembelajaran pada optimizer Adam	0,00025
Loss Function	Fungsi kerugian yang digunakan pada regresi	Huber Loss
Batch Size	Jumlah sampel per iterasi pelatihan	128
Epoch Maksimum	Jumlah iterasi maksimum pada pelatihan	40
MAXLEN (Panjang Urutan)	Panjang maksimum sekuens per sesi pengguna	50 timestep
Dropout	Nilai dropout pada Bi-LSTM layer 1 & 2	0.1 dan 0.1
Callbacks	Teknik stabilisasi pelatihan	EarlyStopping, ReduceLR-OnPlateau

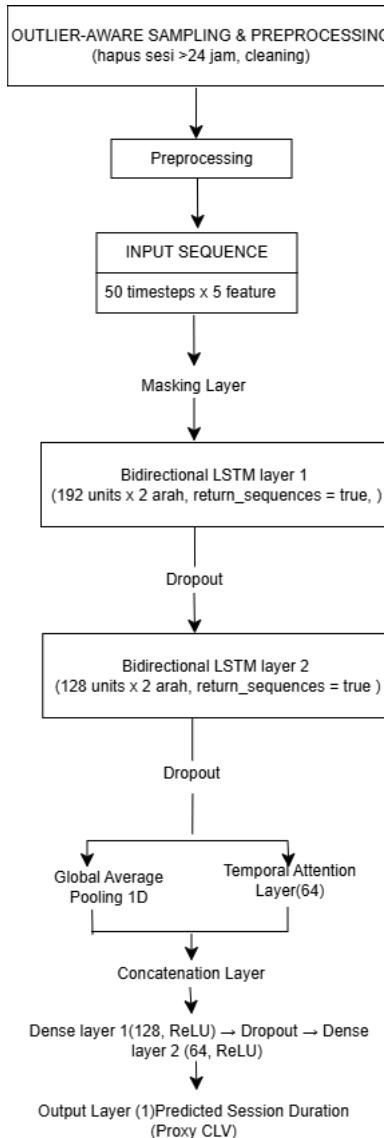
TABEL IX
ARSITEKTUR DAN MEKANISME TEMPORAL ATTENTION LAYER PADA MODEL BI-LSTM V3

Komponen	Deskripsi	Nilai
Input Attention	Vektor keluaran Bi-LSTM layer ke-2	Shape: (batch, 50, 256)
Projection Layer (W)	Proyeksi vektor hidden ke ruang attention	Dense(64, activation= 'tanh')
Score Layer (V)	Mengubah vektor proyeksi menjadi skor attention per timestep	Dense(1, activation= None)
Softmax	Normalisasi skor attention	Softmax sepanjang axis timestep
Masking Attention	Pengabaian timesteps hasil padding	Mask diterapkan
Context Vector	Gabungan bobot × hidden state	Shape: (batch, 256)
Concatenate Output	Penggabungan context vector + GlobalAveragePooling1D	Shape: (batch, 512)

Gambar 4. menunjukkan arsitektur lengkap dari model Modified Bi-LSTM v3 yang digunakan pada penelitian ini. Proses dimulai dari outlier-aware sampling, yang menghapus sesi berdurasi sangat ekstrem (>24 jam) untuk mencegah distorsi distribusi target durasi sesi. Setelah preprocessing, setiap sesi direpresentasikan sebagai input sequence berdimensi 50×5 yang mencakup lima fitur temporal dan perilaku (*event_type_enc*, *category_code_enc*, *brand_enc*, *price*, *time_diff*).

Arsitektur inti terdiri dari dua lapisan Bi-LSTM, masing-masing dengan 192 unit (layer pertama) dan 128 unit (layer kedua). Kedua lapisan ini dikonfigurasi menggunakan *return_sequences=True*, sehingga model dapat menangkap

dependensi jangka panjang dari kedua arah (*forward* dan *backward*) pada seluruh posisi *timestep*. *Dropout* sebesar 0.1 diterapkan pada kedua lapisan untuk mengurangi risiko *overfitting*.



Gambar 4 Arsitektur model Modified Bi-LSTM v3 dengan Temporal Attention Layer dan Outlier-Aware Sampling.

Output dari lapisan Bi-LSTM kedua kemudian diproses melalui *Temporal Attention Layer* (64 unit), yang menghasilkan context vector berbobot, menyoroti event yang paling berpengaruh dalam urutan. Selain itu, *GlobalAveragePooling1D* digunakan untuk merangkum informasi temporal secara komprehensif.

Secara khusus, mekanisme *Attention Layer* ditempatkan setelah lapisan Bi-LSTM kedua yang menghasilkan *return_sequences=True*, sehingga setiap *timestep* memiliki representasi kontekstual lengkap. *Attention* memproyeksikan setiap hidden state melalui Dense (64, tanh) dan menghitung

skor perhatian dengan Dense (1) yang kemudian dinormalisasi menggunakan softmax untuk menentukan bobot tiap timestep. Model dilatih menggunakan *Huber Loss* yang lebih *robust* terhadap *outlier*, dengan *optimizer Adam* dan *learning rate* 0.00025. Untuk menjaga stabilitas pelatihan, diterapkan *EarlyStopping* dan *ReduceLROnPlateau*, sementara data dibagi menggunakan skema *hold-out* 60:20:20 untuk memastikan validitas evaluasi.

G. Evaluasi Model dan Penerapan XAI

Evaluasi dilakukan pada dua pendekatan, yaitu evaluasi prediktif dan evaluasi interpretabilitas. Evaluasi prediktif menggunakan tiga metrik regresi umum, yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE) yang dihitung pada skala 0-1 maupun detik asli untuk memberikan implikasi praktis yang dapat dipahami oleh manajer bisnis.. Rumus masing-masing metrik ditunjukkan pada Persamaan (1)–(3).

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

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

$$RMSE = \sqrt{\sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3)$$

Kedua, interpretabilitas model dilakukan dengan XAI [21]. Penelitian ini menggunakan SHAP untuk mengidentifikasi kontribusi fitur secara global dan LIME untuk menjelaskan prediksi pada level *instance* [22], [23]. Kedua metode ini diakui sebagai teknik XAI yang banyak diterapkan pada sistem AI yang berhadapan dengan konsumen [17], sehingga relevan untuk mengatasi isu *black-box* pada model DL.

III. HASIL DAN PEMBAHASAN

A. Validasi Awal

Tahap awal penelitian difokuskan pada replikasi model yang dikembangkan oleh [3] guna memastikan konsistensi hasil serta membangun dasar pembanding yang kredibel. Dataset yang digunakan adalah REES46 (April 2020), yang memuat 1,6 juta catatan perilaku pengguna dan 276.779 sesi unik. Perbedaan utama dibandingkan penelitian sebelumnya terletak pada metode pengambilan sampel. Studi [3] menggunakan 100.000 baris pertama secara berurutan, sedangkan penelitian ini menerapkan random sampling terhadap 100.000 sesi unik. Pendekatan acak tersebut dimaksudkan untuk memperluas keragaman temporal sehingga model dapat merepresentasikan variasi aktivitas pengguna di berbagai periode waktu. Setelah proses sampling, data kembali diurutkan berdasarkan variabel

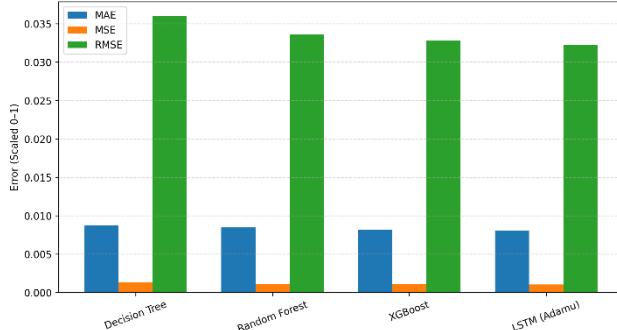
event_time untuk menjaga urutan kronologis yang relevan terhadap konteks temporal.

Sebagai langkah verifikasi metodologis, penelitian ini menggunakan empat model baseline yang mencakup baik pendekatan sekuensial maupun non-sekuensial. Dua baseline awal *Decision Tree* dan LSTM yang di replikasi dari penelitian [3], sedangkan dua model tambahan yaitu *Random Forest* dan XGBoost diikutsertakan untuk menyediakan pembanding berbasis machine learning klasik yang umum digunakan dalam regresi pada data tabular. Keberadaan keempat baseline ini memungkinkan evaluasi performa yang lebih komprehensif, sehingga analisis tidak hanya menitikberatkan pada superioritas arsitektur deep learning, tetapi juga menguji sejauh mana model sekuensial memberikan value added dibandingkan pendekatan prediktif konvensional.

Hasil evaluasi model baseline ditunjukkan pada Tabel 10, sedangkan visualisasi komparatifnya disajikan pada Gambar 5.

TABEL X
HASIL EVALUASI MODEL BASELINE PEMBANDING

Model	Arsitektur	MAE	MSE	RMSE
Decision Tree	Non-Sekuensial	0,0088	0,0013	0,0360
Random Forest	Ensemble (tabular)	0,0085	0,0011	0,0336
XGBoost	Gradient Boosting	0,0081	0,0011	0,0329
LSTM	Konvensional	0,0080	0,0010	0,0322



Gambar 5 Visualisasi Model Pembanding

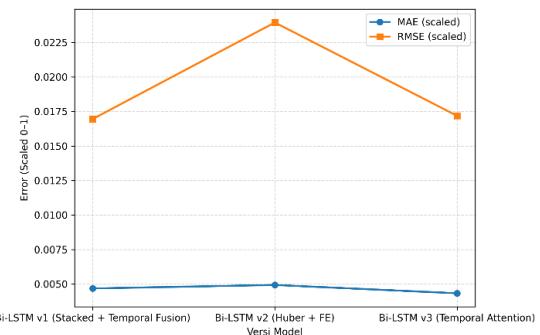
Secara umum, hasil menunjukkan bahwa model LSTM menghasilkan nilai MAE = 0,0080 dan RMSE = 0,0322, lebih baik dibandingkan Decision Tree (MAE = 0,0088). Model ensemble seperti Random Forest dan XGBoost menunjukkan performa kompetitif dengan nilai kesalahan yang mendekati LSTM. Temuan ini mengonfirmasi bahwa pemanfaatan struktur urutan temporal pada LSTM memberikan keunggulan dibandingkan model berbasis aturan statis. Perbedaan kecil terhadap hasil [3] disebabkan oleh variasi distribusi temporal akibat pendekatan sampling acak, namun secara keseluruhan tren performa tetap konsisten. Dengan demikian, hasil replikasi dapat dinyatakan valid dan layak dijadikan dasar untuk pengembangan arsitektur yang lebih kompleks.

B. Model yang Diusulkan: Bi-LSTM v3

Mengacu pada desain arsitektur yang telah dijelaskan pada Bab 2, penelitian ini mengembangkan model utama Modified Bi-LSTM melalui tiga tahap iteratif: v1, v2, dan v3. Setiap versi dirancang secara modular untuk mengevaluasi kontribusi masing-masing komponen terhadap peningkatan performa prediksi durasi sesi sebagai proxy CLV. Pendekatan bertahap ini dirancang untuk mengatasi keterbatasan LSTM konvensional dalam menangkap dependensi dua arah serta dinamika temporal yang kompleks pada data klikstream. Rangkuman hasil eksperimen dari ketiga versi model disajikan pada Tabel 11, sementara tren peningkatan performa divisualisasikan pada Gambar 6.

TABEL XI
HASIL MODIFIED BI-LSTM

Model	MAE	MSE	RMSE	MAE (detik)	RMSE (detik)
V1	0,0047	0,0003	0,0170	433,12	1570,97
V2	0,0049	0,0006	0,024	420,09	2041,65
V3	0,0043	0,0003	0,0172	368,78	1465,61



Gambar 6 Tren peningkatan performa antar versi model Bi-LSTM berdasarkan nilai MAE dan RMSE.

Model v1 (*Stacked + Temporal Fusion*) merupakan versi awal dengan dua lapisan Bi-LSTM berukuran besar untuk menangkap hubungan dua arah dalam urutan event. Versi ini menghasilkan nilai MAE sebesar 0,00467 (≈ 433 detik) dan RMSE sebesar 0,0169 (≈ 1570 detik). Meskipun peningkatan akurasi sudah signifikan dibanding baseline LSTM, performa masih fluktuatif pada sesi berdurasi panjang, terutama pada periode evening berdasarkan analisis day part.

Model v2 (*Huber Loss + Feature Enhancement*) menambahkan fitur statistik dan mengganti loss function menjadi *Huber Loss* untuk meningkatkan ketahanan terhadap outlier. Penggunaan *Huber Loss* berhasil menjaga kestabilan pelatihan dengan nilai MAE sebesar 0,00493 (≈ 420 detik) dan RMSE sebesar 0,0239 (≈ 2041 detik). Analisis per *daypart* menunjukkan bahwa model ini memiliki akurasi terbaik pada rentang waktu evening dengan kesalahan rata-rata 373 detik, sedangkan pada sesi sangat panjang (*very long session*), nilai MAE masih relatif tinggi (1405 detik).

Model v3 (*Temporal Attention + Outlier-Aware Sampling*) merupakan versi akhir dengan penambahan lapisan Temporal Attention untuk memberikan bobot dinamis pada event yang

paling berpengaruh terhadap durasi sesi. Hasil pengujian menunjukkan peningkatan performa paling signifikan dengan MAE sebesar 0,00432 (\approx 368 detik) dan RMSE sebesar 0,0171 (\approx 1466 detik). Selain itu, performa model tetap konsisten di seluruh daypart dengan rentang kesalahan antara 345–448 detik, menunjukkan stabilitas temporal yang lebih baik dibanding versi sebelumnya.

Untuk mengevaluasi kontribusi spesifik dari penambahan *Temporal Attention*, dilakukan komparasi langsung antara v2 (tanpa *attention*) dan v3 (dengan *attention*), sebagaimana ditunjukkan pada Tabel 12.

TABEL XII
DAMPAK TEMPORAL ATTENTION TERHADAP PERFORMA MODEL

Model	Arsitektur	MAE (detik)	RMSE (detik)	Perubahan dari v2
v2	Bi-LSTM tanpa Attention	420,09	2041,65	-
v3	Bi-LSTM + Temporal Attention	368,78	1465,61	Peningkatan 12,2% MAE dan 28,2% RMSE

Hasil pada Tabel 10 menunjukkan bahwa penambahan *Temporal Attention* pada model v3 memberikan peningkatan performa yang substansial, khususnya dalam menurunkan kesalahan pada sesi panjang dan mengurangi variabilitas prediksi.

Hasil eksplorasi *attention weight* pada beberapa sesi menunjukkan bahwa mekanisme *Attention* secara konsisten memberikan bobot tertinggi pada event-event yang terjadi menjelang akhir sesi, perubahan kategori produk, serta interaksi dengan produk bernilai tinggi. Pola perubahan bobot ini menunjukkan bahwa *attention* mampu menangkap dinamika temporal yang bermakna, bukan menghasilkan bobot secara acak.

Selain itu, pola attention tersebut selaras dengan karakteristik perilaku pengguna e-commerce, di mana keputusan paling penting cenderung terjadi menjelang akhir sesi. Dengan demikian, mekanisme Temporal Attention tidak hanya meningkatkan akurasi prediktif, tetapi juga memberikan interpretabilitas temporal tambahan yang tidak tersedia pada model v1 dan v2.

Secara keseluruhan, modifikasi yang diterapkan pada tiap versi Bi-LSTM memberikan efek berbeda terhadap kinerja model. v1 memperkuat kapasitas *pembelajaran* sekuensial, v2 menambah stabilitas namun belum signifikan meningkatkan akurasi, sementara v3 menghasilkan peningkatan paling tinggi melalui penerapan *temporal attention* dan *outlier-aware sampling*. Dengan demikian, model v3 dipilih sebagai representasi terbaik untuk tahap interpretasi model menggunakan Explainable AI.

C. Evaluasi dan Hasil Perbandingan

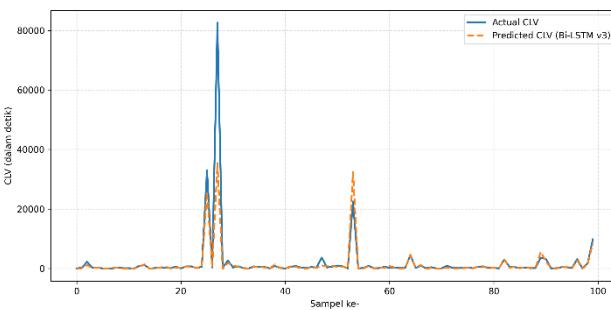
Evaluasi dilakukan menggunakan metrik MAE, MSE, dan RMSE. Tabel 9 dan 10 menunjukkan bahwa model usulan Bi-

LSTM menghasilkan performa yang lebih baik dibandingkan baseline, dengan penurunan nilai kesalahan secara konsisten pada model v3.

Model Bi-LSTM v3 menghasilkan nilai MAE sebesar 0,0043 dan RMSE sebesar 0,0172, yang berarti terjadi peningkatan akurasi sekitar 45–50% dibandingkan LSTM konvensional. Peningkatan tersebut menunjukkan bahwa mekanisme perhatian temporal efektif dalam menyoroti event penting selama sesi pengguna, sehingga prediksi durasi menjadi lebih presisi.

Sementara itu, model XGBoost dan Random Forest menunjukkan performa yang lebih baik dibandingkan Decision Tree, namun masih tertinggal dari pendekatan berbasis urutan waktu. Hal ini menguatkan temuan bahwa model non-sekuensial tidak mampu menangkap konteks kronologis antar-event, sehingga representasi perilaku pengguna menjadi kurang komprehensif.

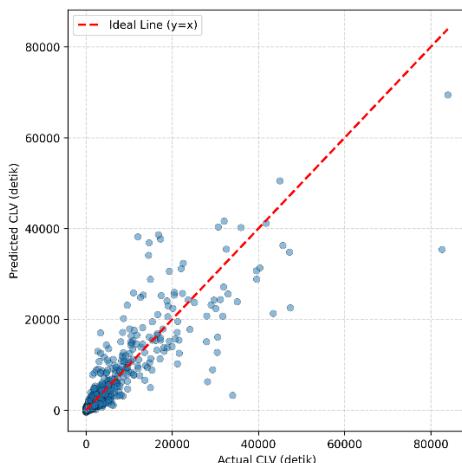
Distribusi hasil prediksi menunjukkan kesesuaian yang tinggi dengan distribusi aktual, terutama pada sesi berdurasi menengah (300–1800 detik). Deviasi meningkat pada sesi sangat panjang (>1800 detik), yang merupakan karakteristik umum dari data perilaku pengguna dengan dinamika waktu tidak linier. Secara keseluruhan, hasil ini membuktikan bahwa model Bi-LSTM v3 mampu merepresentasikan karakteristik temporal dengan tingkat presisi yang tinggi dan menghasilkan performa terbaik di antara seluruh varian yang diuji.



Gambar 7 Perbandingan antara nilai aktual dan prediksi CLV pada data uji menggunakan model Modified Bi-LSTM v3.

Gambar 7. Perbandingan antara nilai aktual dan prediksi CLV pada data uji menggunakan model Modified Bi-LSTM v3. Terlihat bahwa nilai prediksi cenderung mengikuti pola nilai aktual dengan deviasi yang relatif kecil, menunjukkan kemampuan model dalam merepresentasikan hubungan temporal antarperistiwa dengan baik.

Gambar 8. Hubungan antara nilai aktual dan prediksi CLV menggunakan model Modified Bi-LSTM v3. Titik-titik data menunjukkan distribusi hasil prediksi terhadap nilai aktual pada data uji, sedangkan garis merah menandakan prediksi ideal ($y = x$). Sebagian besar titik berada di sekitar garis tersebut, menandakan tingkat kesesuaian yang tinggi dan validitas prediksi yang baik, dengan nilai R^2 sebesar 0.87 yang menunjukkan korelasi kuat antara hasil prediksi dan nilai aktual.



Gambar 8 Hubungan antara nilai aktual dan prediksi CLV menggunakan model Modified Bi-LSTM v3.

D. Analisi Temporal dan Pola Prediksi

Analisis temporal dilakukan untuk mengidentifikasi variasi tingkat kesalahan model terhadap dimensi waktu dan panjang sesi.

TABEL XIII
ANALISIS TEMPORAL MODEL BI-LSTM V3

Daypart	MAE (detik)
Night	380,32
Morning	377,47
Afternoon	369,50
Evening	456,89

TABEL XIV
ANALISIS MAE BERDASARKAN PANJANG SESI

Panjang Sesi	MAE (detik)
Short	100,49
Medium	196,16
Long	443,41
Very Long	2726,86

Tabel 13. menunjukkan bahwa performa model relatif stabil di seluruh rentang waktu aktivitas pengguna, dengan nilai MAE berkisar antara 369–456 detik. Nilai kesalahan terendah tercatat pada periode afternoon (369 detik), yang mengindikasikan bahwa model lebih akurat saat memprediksi pola sesi yang terjadi pada jam aktif pengguna.

Berdasarkan panjang sesi (Tabel 14), model menunjukkan peningkatan kesalahan seiring bertambahnya durasi sesi. Pada sesi short dan medium, nilai MAE tetap rendah (100–196 detik), namun meningkat tajam pada sesi very long (>1800 detik). Hal ini wajar karena sesi panjang umumnya mengandung variasi perilaku yang lebih kompleks dan rentang waktu antar-event yang lebih tidak teratur.

Secara keseluruhan, hasil ini menunjukkan bahwa model Bi-LSTM v3 memiliki generalisasi yang baik terhadap pola temporal harian dan mampu mempertahankan stabilitas prediksi pada mayoritas kelompok sesi pengguna.

E. Interpretasi Bisnis

Interpretasi nilai MAE dan RMSE dilakukan dengan membandingkannya terhadap distribusi durasi sesi aktual. Dengan median durasi sesi sebesar 185 detik dan rata-rata 412 detik, nilai MAE model pada rentang 369–456 detik menunjukkan bahwa kesalahan absolut masih relatif besar pada skala per-sesi. Namun, pada segmen yang mendominasi dataset yaitu sesi pendek dan menengah nilai MAE yang dicapai berada pada rentang 100–196 detik, sehingga model tetap relevan untuk memodelkan perilaku mayoritas pengguna.

Dalam konteks bisnis, durasi sesi digunakan sebagai proxy CLV jangka pendek, terutama untuk analisis engagement dan segmentasi. Kesalahan pada rentang 100–200 detik masih berada dalam batas toleransi untuk aplikasi strategis agregat seperti analisis pola interaksi dan pengelompokan perilaku, meskipun kurang sesuai untuk keputusan mikro-level yang membutuhkan estimasi durasi presisi.

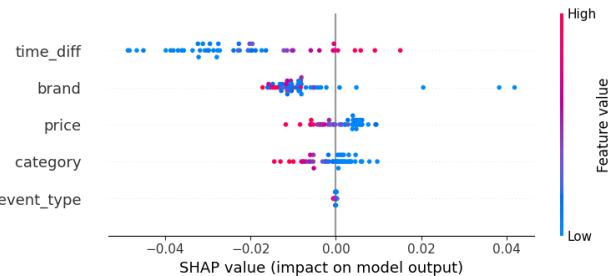
Secara metodologis, model Bi-LSTM v3 menunjukkan peningkatan performa dibandingkan baseline konvensional seperti *Decision Tree*, *Random Forest*, dan *XGBoost*. Hal ini mengindikasikan bahwa meskipun error absolutnya belum rendah, pendekatan sekuensial tetap memberikan nilai tambah relatif dalam mempelajari dinamika temporal dibandingkan metode non-sekuensial.

Dengan demikian, model tetap memiliki utilitas praktis dalam konteks analitik *behavior-based* CLV pada tingkat agregat, sambil menunjukkan keunggulan prediktif yang konsisten atas baseline klasik.

F. Analisis Interpretabilitas Model

Interpretabilitas diperlukan untuk memahami bagaimana model menghasilkan keputusan prediksi dan sejauh mana setiap fitur berkontribusi terhadap hasil tersebut. Untuk itu, analisis XAI diterapkan menggunakan dua pendekatan: SHAP untuk interpretasi global, dan LIME untuk interpretasi lokal. Kedua metode ini digunakan untuk mengevaluasi model Modified Bi-LSTM v3 yang terbukti memiliki performa terbaik pada tahap sebelumnya. Selain memberikan gambaran umum, analisis XAI juga digunakan untuk memvalidasi konsistensi perilaku model dan memastikan interpretasi global serta lokal menunjukkan pola yang selaras.

a. Interpretasi Global dengan SHAP



Gambar 9 Visualisasi SHAP summary plot

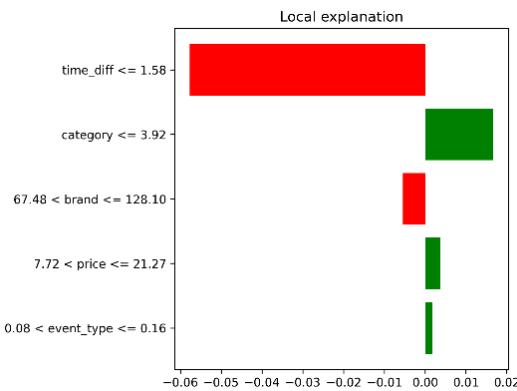
Hasil visualisasi SHAP summary plot (Gambar 9) menunjukkan tingkat kontribusi relatif setiap fitur terhadap keluaran model. Fitur *time_diff* memiliki pengaruh paling dominan terhadap prediksi durasi sesi pengguna, diikuti oleh *brand* dan *price*. Nilai SHAP positif pada fitur *time_diff* menunjukkan bahwa peningkatan selisih waktu antar-event cenderung meningkatkan estimasi durasi total sesi, sedangkan nilai negatif menandakan sesi yang lebih singkat.

Sementara itu, fitur *category* dan *event_type* memiliki pengaruh yang lebih rendah dan bervariasi antar-sampel, mengindikasikan bahwa konteks perilaku pengguna (misalnya jenis aktivitas atau kategori produk) hanya berperan sekunder dibandingkan pola temporal.

Dominasi fitur temporal pada hasil SHAP menguatkan kesimpulan bahwa model Bi-LSTM v3 berhasil mempelajari dinamika waktu sebagai faktor utama dalam prediksi CLV.

Untuk memeriksa konsistensi, dilakukan validasi silang dengan LIME. Hasil SHAP menunjukkan bahwa *time_diff*, *price*, dan *category_code_enc* selalu menempati tiga fitur teratas pada sebagian besar sampel, sehingga membentuk dasar interpretasi global yang kuat.

b. Interpretasi Lokal dengan LIME



Gambar 10 Visualisasi LIME

Untuk memahami keputusan model pada tingkat individu, dilakukan interpretasi lokal menggunakan metode LIME terhadap salah satu sampel pada data uji (Gambar 10). Hasil interpretasi menunjukkan bahwa fitur *price* dan *category* berkontribusi positif terhadap prediksi, artinya nilai harga yang lebih tinggi serta kategori produk tertentu meningkatkan estimasi durasi sesi. Sebaliknya, *time_diff* dengan nilai rendah memberikan kontribusi negatif terhadap hasil prediksi, mengindikasikan sesi yang relatif singkat.

Visualisasi LIME ini memperlihatkan bagaimana kombinasi nilai fitur membentuk keputusan spesifik model terhadap satu pengguna, yang secara keseluruhan konsisten dengan hasil interpretasi global SHAP.

Pengamatan pada beberapa sampel uji lainnya memperlihatkan kecenderungan pola yang serupa, di mana *time_diff* tetap muncul sebagai faktor lokal yang berpengaruh kuat—konsisten dengan hasil analisis global SHAP.

Konsistensi pola kontribusi antara SHAP dan LIME mengindikasikan bahwa model memiliki stabilitas interpretatif dan tidak menunjukkan perilaku kontradiktif antara level global dan individual. Temuan ini juga mendukung desain arsitektur v3, di mana *Temporal Attention* secara efektif menyoroti event penting, sementara *outlier-aware sampling* membantu menstabilkan kontribusi fitur harga.

G. Sintesis Hasil

Secara keseluruhan, hasil eksperimen memperlihatkan bahwa:

1. Replikasi model baseline menghasilkan pola performa yang konsisten dengan penelitian terdahulu, sehingga valid untuk dijadikan pembanding.
2. Arsitektur Modified Bi-LSTM v3 menunjukkan peningkatan performa signifikan dibanding model baseline maupun ensemble tradisional.
3. Analisis XAI menunjukkan bahwa dinamika waktu antar interaksi pengguna menjadi faktor dominan yang memengaruhi durasi sesi, menggantikan variabel statis seperti waktu mulai aktivitas. Temuan dari SHAP menunjukkan bahwa fitur *time_diff* memiliki kontribusi paling konsisten secara global, sementara LIME menyoroti variasi lokal di mana fitur seperti *price* dan *category* dapat mendominasi dalam konteks tertentu. Perbedaan ini menunjukkan bahwa SHAP dan LIME saling melengkapi dalam menjelaskan perilaku model secara menyeluruh baik dari sudut pandang global maupun individual.

Keseluruhan hasil ini menegaskan bahwa integrasi mekanisme temporal attention dan bidirectional processing merupakan pendekatan efektif dalam prediksi nilai pelanggan berbasis urutan perilaku pengguna di lingkungan e-commerce multi-kategori.

IV. KESIMPULAN

Berdasarkan hasil pengujian dan analisis yang telah dilakukan, dapat disimpulkan bahwa seluruh tahapan replikasi dan modifikasi model berhasil menunjukkan peningkatan performa yang konsisten. Replikasi terhadap penelitian [3] yang digunakan sebagai model *baseline* menghasilkan nilai MAE sebesar 0,0080 dan RMSE sebesar 0,0322, yang menjadi acuan untuk pengujian model usulan.

Penerapan tiga varian model Bi-LSTM yang diusulkan (v1, v2, dan v3) menunjukkan peningkatan akurasi bertahap. Model Bi-LSTM v3, yang mengintegrasikan *temporal attention* dan *outlier-aware sampling*, memberikan performa terbaik dengan nilai MAE sebesar 0,0043 dan RMSE sebesar 0,0172 atau peningkatan akurasi sekitar 45–50% dibandingkan LSTM konvensional. Hasil ini membuktikan bahwa mekanisme perhatian temporal efektif dalam menangkap pola kronologis perilaku pengguna dan meningkatkan presisi prediksi durasi sesi. Peningkatan ini juga konsisten ketika dibandingkan dengan baseline non-sekuensial seperti *Decision Tree*, *Random Forest*, dan

XGBoost, sehingga mengonfirmasi keunggulan pendekatan sekuensial berbasis temporal.

Kontribusi ilmiah utama penelitian ini terletak pada integrasi temporal attention dan strategi outlier-aware sampling, yang secara empiris terbukti meningkatkan stabilitas dan akurasi prediksi durasi sesi. Analisis temporal memperlihatkan bahwa model tetap stabil pada seluruh daypart dengan kisaran MAE 369–456 detik, meskipun kesalahan meningkat pada sesi sangat panjang (>1800 detik) akibat kompleksitas perilaku waktu.

Pada tahap interpretabilitas, hasil analisis SHAP dan LIME mengonfirmasi bahwa fitur *time_diff*, *price*, dan *brand* merupakan determinan utama dalam pembentukan prediksi model. Konsistensi antara interpretasi global dan lokal menunjukkan bahwa model tidak hanya unggul secara kuantitatif, tetapi juga memiliki transparansi dalam proses pengambilan keputusan (*interpretable deep learning*).

Secara keseluruhan, penelitian ini berhasil mengembangkan model prediksi berbasis Bi-LSTM dengan perhatian temporal yang mampu merepresentasikan dinamika perilaku pengguna secara akurat dan dapat dijelaskan, menjadikannya solusi potensial untuk analisis nilai pelanggan CLV pada sistem e-commerce modern.

Namun demikian, penelitian ini memiliki beberapa keterbatasan, terutama tingginya MAE pada sesi sangat panjang, penggunaan satu sumber data (REES46), dan belum dievaluasinya model dalam konteks operasional real-time. Penelitian selanjutnya dapat mengeksplorasi arsitektur berbasis Transformer, pemodelan per kategori atau segmen pengguna, serta validasi pada data multi-platform untuk meningkatkan generalisasi.

DAFTAR PUSTAKA

- [1] E. B. Firmansyah, M. R. Machado, and J. L. R. Moreira, "How can Artificial Intelligence (AI) be used to manage Customer Lifetime Value (CLV)—A systematic literature review," *Int. J. Inf. Manag. Data Insights*, vol. 4, no. 2, p. 100279, Nov. 2024, doi: 10.1016/j.jjimei.2024.100279.
- [2] D. Egorenkov, "AI-Powered Predictive Customer Lifetime Value: Maximizing Long-Term Profits," *Int. J. Sci. Res. Manag.*, vol. 12, no. 09, pp. 7339–7354, 2024, doi: 10.18535/ijsrn/v12i09.em02.
- [3] S. Adamu, A. Iorlami, and Ö. Asilkan, "Exploring Explainability in Multi-Category Electronic Markets: A Comparison of Machine Learning and Deep Learning Approaches," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 4, pp. 440–454, Mar. 2025, doi: 10.62411/faith.3048-3719-58.
- [4] G. Yilmaz Benk, B. Badur, and S. Mardikyan, "A New 360° Framework to Predict Customer Lifetime Value for Multi-Category E-Commerce Companies Using a Multi-Output Deep Neural Network and Explainable Artificial Intelligence," *Inf.*, vol. 13, no. 8, 2022, doi: 10.3390/info13080373.
- [5] J. Bauer and D. Jannach, "Improved Customer Lifetime Value Prediction with Sequence-To-Sequence Learning and Feature-Based Models," *ACM Trans. Knowl. Discov. Data*, vol. 15, no. 5, 2021, doi: 10.1145/3441444.
- [6] D. Kvíčala, M. Králová, and P. Suchánek, "The impact of online purchase behaviour on customer lifetime value," *J. Mark. Anal.*, no. 2000, 2024, doi: 10.1057/s41270-024-00328-9.
- [7] A. A. TOKUC and T. Dag, "Predicting User Purchases From Clickstream Data: A Comparative Analysis of Clickstream Data Representations and Machine Learning Models," vol. 13, no. March, 2025, doi: 10.1109/ACCESS.2025.3548267.
- [8] A. Prashanthan, "An Integrated Framework for Optimizing Customer Retention Budget using Clustering, Classification, and Mathematical Optimization," *J. Comput. Theor. Appl.*, vol. 3, no. 1, pp. 45–63, 2025, doi: 10.62411/jcta.13194.
- [9] V. Sakalauskas and D. Kriksciuniene, "Personalized Advertising in E-Commerce: Using Clickstream Data to Target High-Value Customers," *Algorithms*, vol. 17, no. 1, 2024, doi: 10.3390/a17010027.
- [10] W. Ullah, A. Ullah, I. U. Haq, K. Muhammad, M. Sajjad, and S. W. Baik, "CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks," *Multimed. Tools Appl.*, vol. 80, no. 11, pp. 16979–16995, 2021, doi: 10.1007/s11042-020-09406-3.
- [11] D. Anitha Kumari, M. Shuaib Siddiqui, R. Dorbala, R. Megala, K. T. Vigneswara Rao, and N. Srikanth Reddy, "Deep Learning Models for Customer Lifetime Value Prediction in E-commerce," in *2024 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCT)*, IEEE, Apr. 2024, pp. 227–232, doi: 10.1109/ICRTCT61793.2024.10578372.
- [12] A. R. H. Dwika and D. Avianto, "Implementasi Algoritma LSTM untuk Prediksi Harga Cabai Merah Keriting di Yogyakarta," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 1, pp. 635–648, 2024, doi: 10.35870/jimik.v5i1.534.
- [13] I. Malashin, V. Tynchenko, A. Gantimurov, V. Nelyub, and A. Borodulin, "Applications of Long Short-Term Memory (LSTM) Networks in Polymeric Sciences: A Review," *Polymers (Basel.)*, vol. 16, no. 18, pp. 1–44, 2024, doi: 10.3390/polym16182607.
- [14] T. Yang, Y. Cheng, Y. Ren, Y. Lou, M. Wei, and H. Xin, "A Deep Learning Framework for Sequence Mining with Bidirectional LSTM and Multi-Scale Attention," *Proc. 2025 2nd Int. Conf. Innov. Manag. Inf. Syst. ICIS 2025*, pp. 472–476, 2025, doi: 10.1145/3745676.3745751.
- [15] D. R. I. M. Setiadi, A. R. Muslih, S. W. Iriananda, W. Warto, J. Gondohanindijo, and A. A. Ojugo, "Outlier Detection Using Gaussian Mixture Model Clustering to Optimize XGBoost for Credit Approval Prediction," *J. Comput. Theor. Appl.*, vol. 2, no. 2, pp. 244–255, 2024, doi: 10.62411/jcta.11638.
- [16] E. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 32, no. 11, pp. 4793–4813, 2021, doi: 10.1109/TNNLS.2020.3027314.
- [17] M. Sarkar, "Explainable AI In E-Commerce: Enhancing Trust And Transparency In AI-Driven Decisions," *Innov. Eng. J.*, vol. 2, no. 01, pp. 12–39, Jan. 2024, doi: 10.70937/itej.v2i01.53.
- [18] A. Prashanthan, R. Roshan, and M. Maduranga, "RetenNet: A Deployable Machine Learning Pipeline with Explainable AI and Prescriptive Optimization for Customer Churn Management," *J. Futur. Artif. Intell. Technol.*, vol. 2, no. 2, pp. 182–201, 2025, doi: 10.62411/faith.3048-3719-110.
- [19] B. P. Chamberlain, Á. Cardoso, C. H. Bryan Liu, R. Pagliari, and M. P. Deisenroth, "Customer lifetime value prediction using embeddings," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. Part F1296, pp. 1753–1762, 2017, doi: 10.1145/3097983.3098123.
- [20] M. G. S. P. Kusuma *et al.*, "Integrating Quantum, Deep, and Classic Features with Attention-Guided AdaBoost for Medical Risk Prediction," *J. Comput. Theor. Appl.*, vol. 3, no. 2, pp. 171–189, Oct. 2025, doi: 10.62411/jcta.14873.
- [21] A. M. Salih *et al.*, "A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME," *Adv. Intell. Syst.*, 2024, doi: 10.1002/aisy.202400304.
- [22] T. R. Noviandy, G. M. Idroes, and I. Hardi, "An Interpretable Machine Learning Strategy for Antimalarial Drug Discovery with LightGBM and SHAP," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 84–95, 2024, doi: 10.62411/faith.2024-16.
- [23] B. I. Igoche, O. Matthew, P. Bednar, and A. Gegov, "Integrating Structural Causal Model Ontologies with LIME for Fair Machine Learning Explanations in Educational Admissions," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 65–85, 2024, doi: 10.62411/jcta.10501.

