

Implementasi Vanilla LSTM untuk Prediksi Lintasan Kapal Berbasis AIS

I Wayan Sudiarsa^{*1}, Gede Agus Santiago², I Gusti Ayu Anom³, Vanny Indah Febrianti⁴

^{1,2,3,4}Institut Bisnis dan Teknologi Indonesia

e-mail: ¹sudiarsa@instiki.ac.id, ²santiago@instiki.ac.id, ³anomiga@instiki.ac.id,
⁴vanny.febrianti.vf@gmail.com

Abstract – Automatic Identification System (AIS) data provides important spatiotemporal information for vessel monitoring and trajectory prediction. However, AIS-based prediction remains challenging because vessel movement data contain heterogeneous navigation phases, including sailing, stopping, sharp turning, docking, and route reversal. This study evaluates Vanilla Long Short-Term Memory (LSTM) as a baseline deep learning model for one-step-ahead vessel position prediction using AIS data from Bali waters in January 2026. The initial dataset consisted of 47,966 AIS records and was reduced to 47,697 records after basic preprocessing, including validity filtering, AIS message type selection, duplicate removal, coordinate validation, and speed filtering. Two experimental scenarios were designed: Scenario A used all valid trajectory segments, while Scenario B applied trajectory eligibility filtering based on movement and kinematic criteria. The model used five historical AIS points to predict delta longitude and delta latitude, which were then reconstructed into geographic coordinates. Model performance was compared with a persistence baseline using Haversine error as the main metric. The results show that in Scenario A, the persistence baseline outperformed Vanilla LSTM. In contrast, in Scenario B, Vanilla LSTM reduced the mean Haversine error from 1,884.41 m to 1,420.56 m and the median error from 1,635.99 m to 1,291.22 m. These findings indicate that trajectory eligibility filtering is important for improving LSTM-based vessel trajectory prediction.

Keyword - AIS, Galat Haversine, Persistence Baseline, Prediksi Lintasan Kapal, Trajectory Eligibility Filtering, Vanilla LSTM.

Abstrak—Data Automatic Identification System (AIS) menyediakan informasi spasial-temporal yang penting untuk pemantauan kapal dan prediksi lintasan. Namun, prediksi berbasis AIS masih memiliki tantangan karena data pergerakan kapal mengandung fase navigasi yang heterogen, seperti berlayar, berhenti, berbelok tajam, sandar, dan kembali ke arah asal. Penelitian ini mengevaluasi Vanilla Long Short-Term Memory (LSTM) sebagai model baseline deep learning untuk prediksi posisi kapal satu langkah ke depan menggunakan data AIS perairan Bali periode Januari 2026. Dataset awal terdiri atas 47.966 record AIS dan menjadi 47.697 record setelah praproses dasar, meliputi penyaringan validitas, pemilihan tipe pesan AIS posisi kapal, penghapusan duplikasi, validasi koordinat, dan penyaringan kecepatan. Dua skenario eksperimen dirancang, yaitu Skenario A yang menggunakan seluruh segmen lintasan valid dan Skenario B yang menerapkan trajectory eligibility filtering berdasarkan kriteria gerak dan kinematik. Model menggunakan lima titik AIS historis untuk memprediksi delta longitude dan delta latitude, kemudian direkonstruksi menjadi koordinat geografis. Kinerja model dibandingkan dengan persistence baseline menggunakan galat Haversine sebagai metrik utama. Hasil menunjukkan bahwa pada Skenario A, persistence baseline lebih unggul daripada Vanilla LSTM. Sebaliknya, pada Skenario B, Vanilla LSTM menurunkan rerata galat Haversine dari 1.884,41 m menjadi 1.420,56 m dan median galat dari 1.635,99 m menjadi 1.291,22 m. Temuan ini menunjukkan bahwa trajectory eligibility filtering penting untuk meningkatkan prediksi lintasan kapal berbasis LSTM.

Kata Kunci – AIS, Galat Haversine, Persistence Baseline, Prediksi Lintasan Kapal, Trajectory Eligibility Filtering, Vanilla LSTM.

I. PENDAHULUAN

Transportasi laut merupakan tulang punggung perdagangan internasional, dengan lebih dari 90% volume kargo global diangkut melalui jalur laut [1]. Di Indonesia, kondisi ini semakin kritis mengingat posisi strategis negara sebagai kepulauan terbesar yang menghubungkan Samudra Hindia dan Pasifik. Berdasarkan laporan Kementerian Perhubungan Republik Indonesia, jumlah armada kapal laut nasional mencapai 72.313 unit pada tahun 2021, meningkat 13,9% dibanding tahun sebelumnya [2]. Sejalan dengan pertumbuhan armada tersebut, data dari Direktorat Jenderal Kesatuan Penjagaan Laut dan Pantai mencatat tren peningkatan kecelakaan kapal selama tiga tahun berturut-turut: 87 kejadian (2020), 100 kejadian (2021), dan 108 kejadian (2022) [3]. Kondisi ini mempertegas kebutuhan terhadap sistem pemantauan dan prediksi lintasan kapal yang lebih cerdas dan berbasis data.

Automatic Identification System (AIS) merupakan teknologi pelacakan maritim yang secara periodik memancarkan informasi posisi, waktu, kecepatan (Speed Over Ground/SOG), dan arah gerak (Course Over Ground/COG) kapal. Data AIS dimanfaatkan secara luas untuk pemantauan kepadatan lalu lintas laut, deteksi anomali, serta keselamatan navigasi [4], [5]. Prediksi lintasan kapal (vessel trajectory prediction) merupakan salah satu aplikasi strategis dari data AIS, karena posisi kapal di masa depan menjadi prasyarat deteksi risiko tabrakan secara aktif [1]. Namun, data AIS memiliki karakteristik yang kompleks: interval pelaporan tidak seragam, rentan terhadap duplikasi dan lompatan spasial, serta mencakup berbagai fase gerak kapal seperti berlayar, sandar, memutar haluan (U-turn), dan kembali ke titik asal. Karakteristik heterogen ini menimbulkan tantangan tersendiri dalam persiapan data untuk pemodelan prediktif berbasis deep learning [7], [8].

Long Short-Term Memory (LSTM) telah menjadi pendekatan dominan dalam prediksi lintasan kapal berbasis deret waktu, karena kemampuannya menangkap dependensi temporal jangka panjang pada data sekuensial [1], [6]. Vanilla LSTM bentuk paling dasar dari LSTM yang terdiri atas satu lapisan LSTM tersambung penuh — sering digunakan sebagai model baseline sebelum eksperimen dengan varian yang lebih kompleks [9], [10]. Studi perbandingan arsitektur LSTM pada data AIS perairan Bali mengindikasikan bahwa kedalaman arsitektur berpengaruh terhadap kemampuan representasi fitur [10]. Namun, seluruh studi tersebut menggunakan segmen lintasan valid tanpa mempertimbangkan kelayakan kinematik segmen sebagai data latih. Pertanyaan yang belum terjawab adalah: apakah kinerja Vanilla LSTM dapat ditingkatkan bukan melalui perubahan arsitektur, melainkan melalui penyaringan berbasis kelayakan lintasan sebelum pelatihan?

Sebagian besar penelitian terdahulu mengenai prediksi lintasan kapal berbasis AIS berfokus pada pengembangan arsitektur model, seperti LSTM, GRU, Bi-LSTM, Seq2Seq, attention, maupun model berbasis graf spasial-temporal. Pendekatan tersebut memang penting untuk meningkatkan kemampuan representasi temporal dan spasial, tetapi masih menyisakan celah pada aspek kualitas segmen lintasan sebagai data latih. Data AIS tidak hanya berisi pergerakan kapal yang kontinu, tetapi juga mencakup fase berhenti, sandar, berbelok tajam, memutar arah, dan kembali ke titik asal. Jika seluruh fase tersebut dicampur dalam satu proses pelatihan, model sekuensial dapat menerima pola input yang tidak homogen secara kinematik.

Berdasarkan celah tersebut, penelitian ini tidak berfokus pada pengembangan arsitektur LSTM yang lebih kompleks, melainkan mengevaluasi apakah peningkatan performa dapat diperoleh melalui trajectory eligibility filtering. Kontribusi penelitian ini adalah: (1) menyusun pipeline prediksi posisi kapal berbasis AIS menggunakan Vanilla LSTM; (2) menguji pengaruh filtering kelayakan lintasan terhadap performa prediksi; dan (3) membandingkan Vanilla LSTM dengan persistence baseline untuk menilai nilai tambah model terhadap pendekatan prediksi paling sederhana.

II. PENELITIAN YANG TERKAIT

Pada tabel 1 perbandingan penelitian terdahulu dapat dilihat prediksi lintasan kapal berbasis AIS telah berkembang dari model fisik menuju deep learning. RNN dan variannya khususnya LSTM dan GRU telah menjadi metode dominan karena kemampuannya menangkap relasi deret waktu pada data historis AIS [1]. Wang et al. mengusulkan model SMCHN (Sparse Multi-graph Convolutional Hybrid Network) yang membangun graf spasial dan temporal untuk menangkap interaksi antar kapal secara adaptif, dan menempatkan Vanilla LSTM sebagai baseline yang hanya mempertimbangkan dependensi temporal [1]. Hasilnya menunjukkan bahwa Vanilla LSTM kompetitif pada prediksi jangka pendek, namun performanya menurun pada prediksi jangka panjang akibat akumulasi galat iteratif [1]. Capobianco et al. mengembangkan struktur encoder-decoder berbasis Bi-LSTM dengan mekanisme attention yang memetakan sekuens input ke representasi konteks untuk menghasilkan lintasan prediksi secara sekuensial [6]. Studi-studi ini umumnya berfokus pada pengembangan arsitektur model dan tidak secara eksplisit membahas dampak homogenitas segmen lintasan terhadap kualitas pelatihan.

Sudiarsa et al. melakukan perbandingan empiris empat arsitektur LSTM Vanilla, Stacked, BiLSTM, dan Seq2Seq pada data AIS perairan utara Bali menggunakan galat Haversine sebagai metrik evaluasi [10]. Hasil ini menjadi rujukan langsung bagi penelitian ini dalam memposisikan Vanilla LSTM sebagai baseline dan merancang eksperimen dengan variasi kualitas segmen, bukan variasi arsitektur. Pada sisi infrastruktur data, Sudiarsa et al. menyimpulkan melalui kajian sistematis bahwa arsitektur lakehouse berpotensi mendukung pemrosesan data AIS secara real-time, namun implementasi spesifik untuk kebutuhan maritim masih sangat terbatas [11].

Kualitas data AIS merupakan faktor kritis yang sering diabaikan dalam penelitian prediksi lintasan. Stach et al. menegaskan bahwa data AIS sangat bervariasi kualitasnya tergantung sumber penerima dan kondisi

transmisi, sehingga preprocessing yang tepat menjadi prasyarat sebelum penerapan algoritma cerdas apa pun [5]. Dalam pemrosesan data AIS, interpolasi linear pada celah temporal sering digunakan untuk melengkapi nilai posisi yang hilang, sementara SOG dan heading diisi menggunakan nilai rata-rata dalam periode yang bersangkutan [1]. Penggunaan interpolasi tanpa mempertimbangkan kelas fase gerak kapal berpotensi menghasilkan data latih yang tidak konsisten secara kinematik. Park et al. menunjukkan bahwa heterogenitas tinggi data maritim menjadi hambatan utama pengelolaan data berbasis AI, dan arsitektur vessel data lakehouse terbukti meningkatkan efisiensi pengelolaan data AIS secara signifikan [12]. Permasalahan kualitas data ini diakui secara luas dalam literatur, namun dampaknya terhadap kinerja model prediksi lintasan belum dikaji sebagai variabel eksperimental yang terkontrol.

TABEL I
PERBANDINGAN PENELITIAN TERDAHULU

Penelitian	Fokus	Metode	Celah Penelitian
Wang et al. [1]	Prediksi lintasan berbasis relasi spasial-temporal	SMCHN dan Vanilla LSTM sebagai baseline	Fokus pada arsitektur graf, belum mengevaluasi kelayakan segmen lintasan sebagai variabel eksperimen
Capobianco et al. [6]	Prediksi lintasan berbasis recurrent network	Encoder-decoder Bi-LSTM dan attention	Menekankan peningkatan arsitektur model, bukan homogenitas data latih
Yang et al. [9]	Prediksi lintasan kapal berbasis AIS	Bi-LSTM	Belum membahas pengaruh filtering kinematik segmen terhadap performa prediksi
Sudiarsa et al. [10]	Perbandingan arsitektur LSTM pada data AIS	Vanilla, Stacked, Bi-LSTM, Seq2Seq	Fokus pada variasi arsitektur, belum mengisolasi kualitas segmen lintasan
Penelitian ini	Pengaruh filtering kelayakan lintasan	Vanilla LSTM dan persistence baseline	Mengevaluasi pengaruh kualitas segmen lintasan terhadap performa prediksi tanpa memodifikasi arsitektur

Trajectory eligibility filtering dalam penelitian ini dipahami sebagai proses seleksi segmen lintasan berdasarkan karakteristik gerak kapal agar data latih lebih sesuai untuk prediksi posisi. Konsep ini berbeda dari pembersihan data biasa karena tidak hanya menghapus nilai tidak valid, tetapi juga mendefinisikan segmen yang bergerak, kontinu, dan relatif homogen secara kinematik. Dengan demikian, filtering digunakan untuk membatasi ruang lingkup prediksi pada kondisi lintasan yang lebih layak dipelajari oleh model sekuensial.

Berbeda dari penelitian yang berfokus pada pengembangan arsitektur model baru atau infrastruktur data maritim, penelitian ini menguji secara terkontrol apakah trajectory eligibility filtering — seleksi segmen lintasan berdasarkan kelayakan kinematik — dapat meningkatkan kinerja Vanilla LSTM melebihi persistence baseline tanpa modifikasi arsitektur. Dengan mengisolasi variabel kualitas segmen dari variabel arsitektur, penelitian ini memberikan pembuktian empiris bahwa persiapan data merupakan faktor determinan dalam akurasi prediksi lintasan berbasis AIS, sekaligus menjadi landasan untuk pengembangan pipeline data yang lebih matang pada penelitian lanjutan [13].

III. METODE PENELITIAN

A. Dataset dan Praproses

Dataset yang digunakan merupakan data AIS perairan Bali periode Januari 2026. Data awal terdiri atas 47.966 record dengan 25 kolom, termasuk MMSI, timestamp, longitude, latitude, speed over ground (SOG), course over ground (COG), heading (HDG), rate of turn (ROT), tipe pesan AIS, dan atribut tambahan lain.

TABEL II
PROFIL DATASET DAN RINGKASAN

Metrik	Nilai
Data awal	47.966 record
MMSI unik awal	354
Periode data	1-31 Januari 2026

Data akhir	47.697 record
MMSI unik akhir	353
Data dihapus	269 record

Praproses dilakukan dengan menyaring record valid, mempertahankan tipe pesan posisi kapal yaitu AIS type 1, 2, 3, 18, dan 19, menghapus AIS type 4 karena merepresentasikan base station, menghapus duplikasi MMSI dan timestamp, memvalidasi koordinat, serta menghapus nilai SOG tidak realistis. Nilai HDG = 511 dan ROT = -128 diperlakukan sebagai kode tidak tersedia. Ringkasan dataset dan praproses ditunjukkan pada Tabel I.

Alur penelitian yang digunakan dalam studi ini ditunjukkan pada Gbr. 1. Alur tersebut dimulai dari akuisisi data AIS, praproses dasar, segmentasi lintasan, pembentukan dua skenario eksperimen, feature engineering, pembentukan sequence LSTM, pelatihan Vanilla LSTM, hingga evaluasi menggunakan galat Haversine dan persistence baseline.



Gbr. 1 Alur penelitian prediksi lintasan kapal berbasis AIS

B. Skenario Segmentasi Lintasan

Dua skenario dibangun untuk mengevaluasi pengaruh kelayakan lintasan. Skenario A menggunakan seluruh segmen lintasan AIS valid setelah segmentasi berbasis jeda waktu 60 menit. Segmen dengan jumlah titik kurang dari lookback + 1 tidak digunakan. Skenario ini menggambarkan kondisi umum setelah praproses dasar.

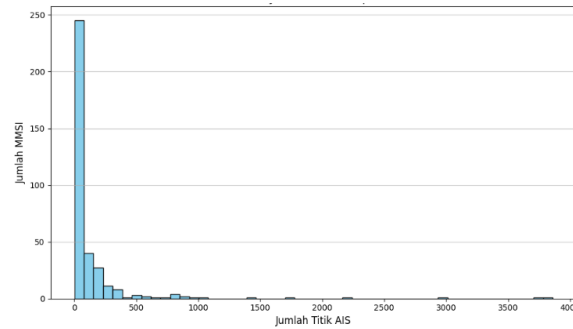
Skenario B menerapkan filtering kelayakan prediksi. Titik dengan SOG kurang dari 0,5 knot dihapus untuk mengurangi dominasi kondisi berhenti atau hampir diam. Segmentasi dilakukan dengan jeda waktu 30 menit dan ambang perubahan COG 60 derajat. Belokan tajam tidak diperlakukan sebagai data buruk, tetapi dijadikan batas segmen agar setiap segmen memiliki karakter gerak yang lebih homogen. Segmen akhir dipertahankan jika memiliki minimal 20 titik, median SOG minimal 1 knot, median jeda waktu maksimal 600 detik, dan durasi segmen lebih dari nol.

TABEL III
RINGKASAN SKENARIO EKSPERIMEN

Skenario	Data	Segmen/Sequence
A	45.103	913 / 40.538
B	5.236	139 / 4.541

Nilai ambang pada Skenario B ditetapkan berdasarkan kombinasi karakteristik domain AIS dan eksplorasi awal dataset. Ambang $SOG \geq 0,5$ knot digunakan untuk mengurangi dominasi titik kapal yang berhenti atau hampir diam. Ambang perubahan COG 60 derajat digunakan sebagai penanda perubahan arah tajam, sehingga titik tersebut dijadikan batas segmen agar lintasan dalam satu segmen lebih homogen. Panjang minimum 20 titik dipilih agar setiap segmen memiliki cukup observasi untuk membentuk beberapa sequence dengan lookback lima titik. Median jeda waktu maksimum 600 detik digunakan untuk mempertahankan segmen yang masih memiliki kontinuitas temporal memadai. Dengan demikian, filtering tidak bertujuan menghapus data yang menghasilkan galat besar, tetapi mendefinisikan ruang lingkup prediksi pada segmen kapal yang bergerak, kontinu, dan layak dipelajari oleh model sekuensial.

Distribusi jumlah titik AIS per MMSI ditampilkan pada Gbr. 2. Gambar tersebut menunjukkan bahwa jumlah titik AIS tidak merata antar-MMSI, dengan sebagian besar kapal memiliki jumlah titik relatif sedikit, sedangkan sebagian kecil MMSI memiliki jumlah titik yang sangat besar.



Gbr. 2 Distribusi jumlah titik AIS per MMSI

C. Feature Engineering dan Model LSTM

Fitur input terdiri atas longitude, latitude, SOG, representasi sinus dan cosinus dari COG, representasi sinus dan cosinus dari HDG, flag ketersediaan HDG dan ROT, delta waktu antar titik, jam, hari dalam minggu, dan hari dalam bulan. Representasi sinus dan cosinus digunakan untuk menghindari masalah diskontinuitas pada variabel sudut.

Model Vanilla LSTM menggunakan lookback lima titik AIS historis. Target tidak berupa koordinat absolut, tetapi perubahan posisi relatif, yaitu delta longitude dan delta latitude terhadap titik input terakhir. Strategi ini digunakan agar model mempelajari pergerakan relatif kapal, bukan hanya skala koordinat absolut. Model terdiri atas satu layer LSTM berisi 20 unit dan satu Dense layer dengan dua neuron output. Optimizer yang digunakan adalah Adam, loss function MSE, batch size 32, dan epoch maksimal 20 dengan EarlyStopping patience 5.

TABEL IV
KONFIGURASI MODEL VANILLA LSTM

Parameter	Nilai
Lookback	5 titik
LSTM unit	20
Output	2 neuron
Split	70:15:15
Baseline	Persistence

D. Baseline dan Metrik Evaluasi

Persistence baseline digunakan sebagai tolok ukur minimum. Baseline ini memprediksi posisi berikutnya sama dengan posisi terakhir pada urutan input, sebagaimana ditunjukkan pada (1). Baseline tidak diposisikan sebagai metode utama kedua, melainkan kontrol evaluasi untuk menilai apakah Vanilla LSTM memberikan nilai tambah dibanding prediksi paling sederhana.

$$\text{lon}(t+1)=\text{lon}(t), \text{lat}(t+1)=\text{lat}(t) \quad (1)$$

Kinerja model dievaluasi menggunakan MAE longitude, MAE latitude, RMSE longitude, RMSE latitude, coordinate RMSE, serta galat Haversine dalam meter. Galat Haversine dipilih karena longitude dan latitude merupakan koordinat geografis sehingga selisih derajat perlu dikonversi menjadi jarak permukaan bumi.

IV. HASIL DAN PEMBAHASAN

A. Hasil Praproses dan Segmentasi

Praproses dasar mengurangi data dari 47.966 menjadi 47.697 record. Pengurangan tersebut terutama berasal dari pembuangan AIS type 4, duplikasi MMSI-timestamp, dan satu record dengan SOG tidak valid. Pada Skenario A, segmentasi menghasilkan 1.990 segmen awal dan 913 segmen valid dengan 40.538 sequence. Pada Skenario B, filtering SOG $\geq 0,5$ knot mengurangi data menjadi 13.603 record, kemudian segmentasi berbasis jeda waktu dan perubahan COG menghasilkan 3.003 segmen. Setelah filtering kelayakan prediksi, tersisa 5.236 record, 139 segmen, dan 4.541 sequence.

Hasil ini menunjukkan bahwa filtering Skenario B jauh lebih ketat dibanding Skenario A. Walaupun jumlah sequence menurun, data yang digunakan menjadi lebih terarah pada lintasan yang bergerak, kontinu, dan relatif homogen. Hal ini penting karena data AIS yang mencampur fase berhenti, berbelok tajam, dan kembali ke arah awal dapat melemahkan proses pembelajaran model sekuensial.

B. Hasil Evaluasi Model

TABEL IV
PERBANDINGAN KINERJA SKENARIO DAN MODEL

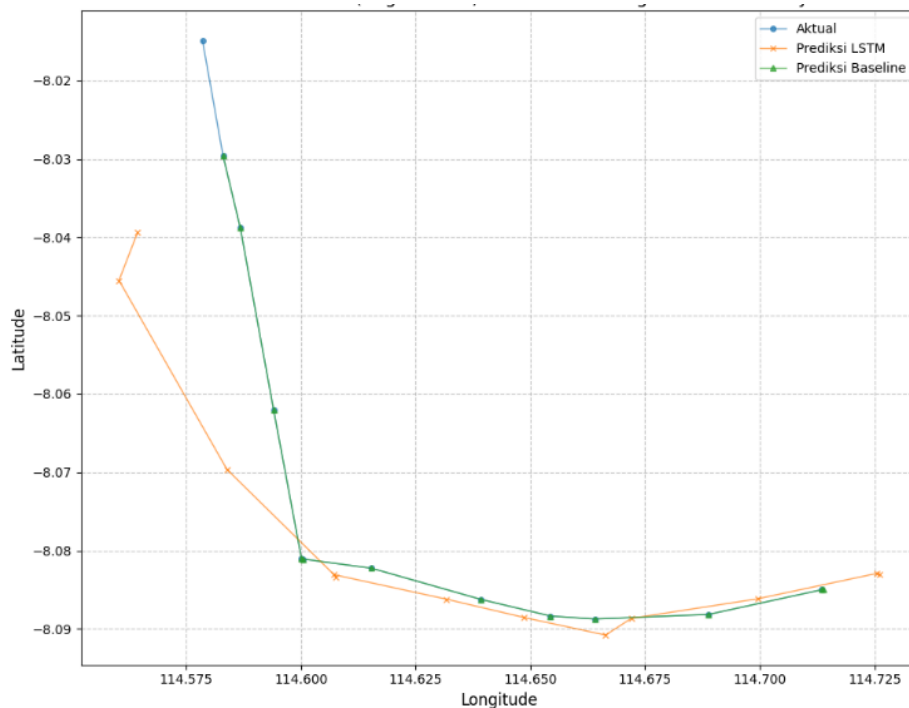
Skenario-Model	Mean H	Median H	P95 H
A-LSTM	758,78	507,74	2.231,99
A-Base	375,03	10,12	2.709,44
B-LSTM	1.420,56	1.291,22	2.887,97
B-Base	1.884,41	1.635,99	4.648,86

Tabel IV menunjukkan bahwa pada Skenario A, persistence baseline masih lebih unggul daripada LSTM berdasarkan rerata dan median galat Haversine. Baseline memperoleh rerata 375,03 m dan median 10,12 m, sedangkan LSTM memperoleh rerata 758,78 m dan median 507,74 m. Hal ini menunjukkan bahwa pada seluruh segmen valid, prediksi satu langkah ke depan sering kali masih sangat dekat dengan posisi terakhir. Dalam kondisi seperti ini, baseline sederhana menjadi sangat kuat.

Keunggulan persistence baseline pada Skenario A dapat dijelaskan oleh sifat prediksi satu langkah ke depan pada data AIS. Pada banyak kasus, posisi AIS berikutnya masih berada sangat dekat dengan posisi terakhir, terutama ketika interval pengiriman data relatif pendek atau kapal berada pada fase bergerak lambat. Dalam kondisi tersebut, prediksi sederhana berbasis posisi terakhir dapat menghasilkan galat yang sangat rendah. Selain itu, Skenario A masih mencampur berbagai fase gerak kapal, termasuk berhenti, berbelok tajam, dan perubahan arah. Heterogenitas ini membuat model LSTM sulit membentuk pola temporal yang konsisten karena urutan historis tidak selalu merepresentasikan arah gerak berikutnya.

Namun, pola berbeda muncul pada Skenario B. Setelah filtering kelayakan lintasan, LSTM menghasilkan rerata galat Haversine 1.420,56 m, lebih rendah daripada baseline sebesar 1.884,41 m. Median galat LSTM sebesar 1.291,22 m juga lebih rendah daripada baseline sebesar 1.635,99 m. Pada P95, LSTM memperoleh 2.887,97 m, sedangkan baseline mencapai 4.648,86 m. Dengan demikian, pada segmen lintasan yang lebih sesuai untuk prediksi, Vanilla LSTM mampu memberikan nilai tambah dibanding baseline.

Untuk memberikan gambaran visual terhadap hasil prediksi, Gbr. 3 menampilkan contoh perbandingan antara lintasan aktual, prediksi Vanilla LSTM, dan prediksi persistence baseline pada salah satu segmen data uji. Visualisasi ini digunakan untuk melihat sejauh mana pola lintasan hasil prediksi mengikuti pola lintasan aktual.

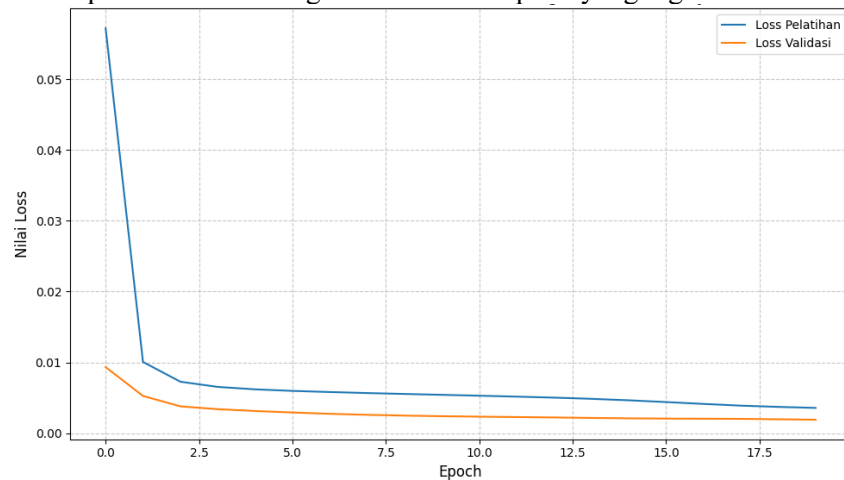


Gbr. 3 Lintasan aktual dan prediksi

Secara relatif, pada Skenario B LSTM menurunkan rerata galat Haversine sekitar 24,62% dibanding baseline. Median galat menurun sekitar 21,08%, sedangkan P95 menurun sekitar 37,88%. Penurunan pada P95 penting karena menunjukkan bahwa filtering kelayakan lintasan tidak hanya memperbaiki nilai rata-rata, tetapi juga menekan galat besar pada sebagian sampel uji.

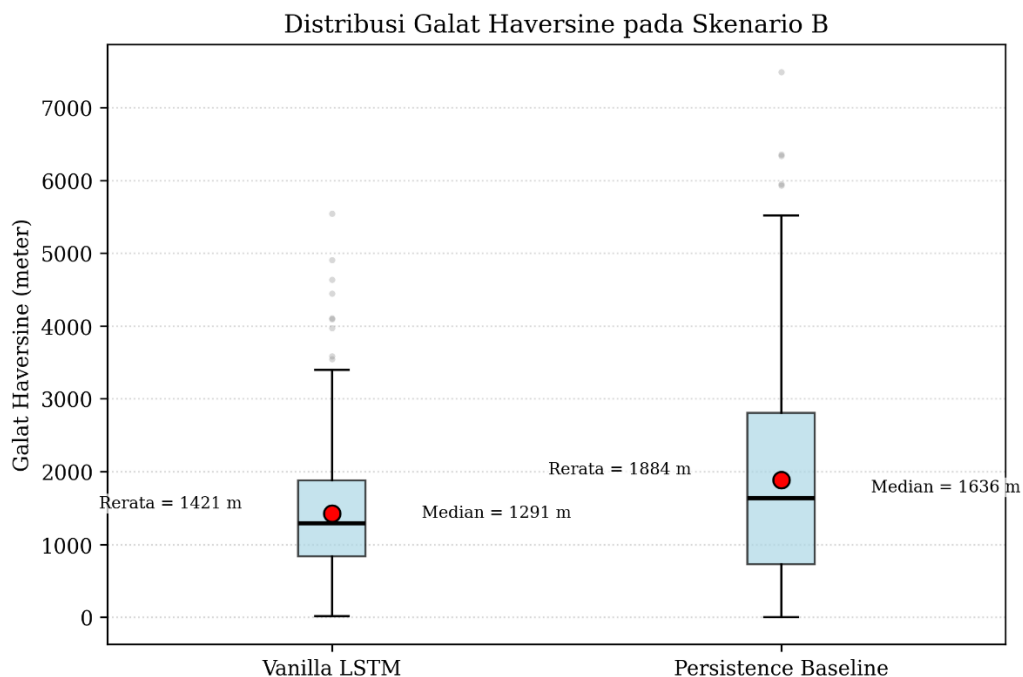
Sebaliknya, peningkatan performa LSTM pada Skenario B menunjukkan bahwa homogenitas segmen berperan penting dalam pembelajaran model. Setelah titik berhenti dikurangi dan perubahan arah tajam diperlakukan sebagai batas segmen, sequence yang digunakan menjadi lebih konsisten secara kinematik. Kondisi ini memungkinkan Vanilla LSTM mempelajari pola perubahan posisi relatif dengan lebih baik dibandingkan baseline yang hanya mempertahankan posisi terakhir. Dengan demikian, hasil penelitian ini menunjukkan bahwa kualitas dan kelayakan segmen lintasan merupakan faktor penting dalam prediksi lintasan kapal berbasis AIS.

Proses pelatihan model pada Skenario B ditunjukkan pada Gbr. 4. Kurva loss pelatihan dan validasi memperlihatkan penurunan nilai loss pada epoch awal dan kemudian cenderung stabil, sehingga menunjukkan bahwa model telah mencapai kondisi konvergen dalam batas epoch yang digunakan.



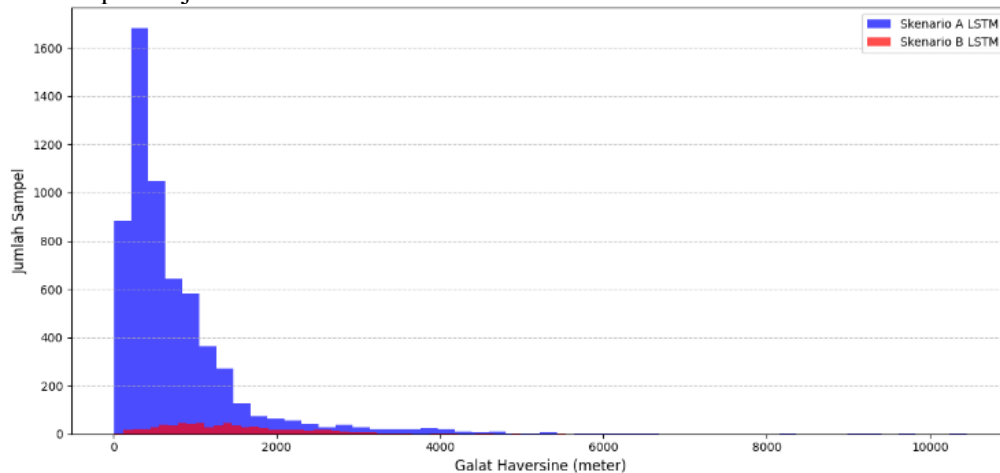
Gbr. 4 Kurva loss pelatihan dan validasi Skenario B

Distribusi galat Haversine antara Vanilla LSTM dan persistence baseline pada Skenario B divisualisasikan pada Gbr. 5. Gambar tersebut memperkuat hasil pada Tabel IV bahwa Vanilla LSTM menghasilkan sebaran galat yang lebih rendah dibandingkan baseline pada segmen lintasan yang telah memenuhi kriteria kelayakan prediksi.



Gbr. 5 Perbandingan galat Haversine LSTM dan baseline

Perbandingan distribusi galat Haversine LSTM antara Skenario A dan Skenario B ditampilkan pada Gbr. 6. Gambar ini menunjukkan perbedaan karakteristik galat antara seluruh segmen valid dan segmen yang telah difilter berdasarkan kelayakan prediksi, sehingga mendukung pembahasan bahwa karakteristik data masukan berpengaruh terhadap kinerja model.



Gbr. 6 grafik density/log scale

C. Implikasi dan Keterbatasan

Temuan utama penelitian ini bukan bahwa Vanilla LSTM selalu unggul, melainkan bahwa kualitas segmen lintasan sangat menentukan efektivitas prediksi AIS. Pada data umum, persistence baseline dapat menjadi pembanding yang sulit dikalahkan karena interval posisi berikutnya sering sangat dekat dengan posisi terakhir. Akan tetapi, ketika data difilter menjadi segmen bergerak dan lebih homogen, LSTM mampu menangkap pola perubahan posisi yang tidak direpresentasikan oleh baseline.

Keterbatasan penelitian ini adalah prediksi dilakukan untuk horizon satu langkah ke depan dan hanya menggunakan Vanilla LSTM. Model belum dibandingkan dengan varian yang lebih kompleks seperti Bi-LSTM, Seq2Seq, attention, atau Transformer. Selain itu, data hanya berasal dari satu periode dan satu wilayah, sehingga generalisasi terhadap wilayah lain masih perlu diuji.

V. KESIMPULAN

Penelitian ini berhasil membangun pipeline prediksi lintasan kapal berbasis AIS menggunakan Vanilla LSTM dengan target delta longitude dan delta latitude. Dataset awal berjumlah 47.966 record dan setelah praproses dasar menjadi 47.697 record.

Pada Skenario A, persistence baseline lebih baik daripada Vanilla LSTM berdasarkan rerata dan median galat Haversine. Hal ini menunjukkan bahwa pada seluruh segmen valid, prediksi satu langkah ke depan sering dapat didekati secara kuat oleh posisi terakhir.

Pada Skenario B, Vanilla LSTM mengungguli baseline dengan rerata galat Haversine 1.420,56 m dibanding 1.884,41 m. Hasil ini menunjukkan bahwa trajectory eligibility filtering penting untuk meningkatkan kualitas prediksi LSTM pada data AIS.

Hasil penelitian ini memiliki implikasi praktis bahwa trajectory eligibility filtering dapat diterapkan sebagai tahap awal dalam pipeline sistem pemantauan dan prediksi lintasan kapal berbasis AIS. Sebelum model prediksi digunakan, data AIS perlu disegmentasi dan diseleksi berdasarkan karakteristik gerak agar model tidak dilatih pada campuran fase navigasi yang tidak homogen. Tahap ini dapat membantu sistem pemantauan maritim menghasilkan prediksi yang lebih stabil pada segmen lintasan yang bergerak dan kontinu.

Penelitian lanjutan dapat diarahkan pada pengujian horizon prediksi multi-langkah, perbandingan dengan model yang lebih kompleks seperti Bi-LSTM, GRU, Seq2Seq, attention, dan Transformer, serta evaluasi lintas wilayah dan lintas musim. Evaluasi lintas wilayah dan lintas musim penting untuk mengetahui kemampuan generalisasi filtering dan model terhadap variasi pola lalu lintas kapal yang berbeda.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada DRPM Institut Bisnis dan Teknologi Indonesia atas dukungan hibah penelitian, serta kepada Konsorsium Penelitian AIS ITS dan Udayana atas dukungan data dan fasilitas penelitian.

DAFTAR PUSTAKA

- [1] S. Wang, Y. Li, Z. Zhang, dan H. Xing, "Big data driven vessel trajectory prediction based on sparse multi-graph convolutional hybrid network with spatio-temporal awareness," *Ocean Engineering*, vol. 287, hal. 115695, 2023. <https://doi.org/10.1016/j.oceaneng.2023.115695>
- [2] M. A. Rizaty, "Ada 72.313 Kapal Laut di Indonesia pada 2021," *DataIndonesia.id*, Sep. 2022. [Online]. Tersedia: <https://dataindonesia.id/otomotif-transportasi/detail/ada-72313-kapal-laut-di-indonesia-pada-2021>
- [3] Redaksi, "Tiga Tahun Terakhir Jumlah Kecelakaan Kapal Laut Meningkat," *Darilaut.id*, Agu. 2023. [Online]. Tersedia: <https://darilaut.id/berita/tiga-tahun-terakhir-jumlah-kecelakaan-kapal-laut-meningkat>
- [4] J. Ye et al., "Deep learning in maritime autonomous surface ships: Current development and challenges," *J. Mar. Sci. Appl.*, 2022.
- [5] T. Stach, Y. Kinkel, M. Constapel, dan H.-C. Burmeister, "Maritime anomaly detection for vessel traffic services: A survey," *J. Mar. Sci. Eng.*, vol. 11, hal. 1174, 2023. <https://doi.org/10.3390/jmse11061174>
- [6] S. Capobianco, L. M. Millefiori, N. Forti, P. Braca, dan P. Willett, "Deep learning methods for vessel trajectory prediction based on recurrent neural networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 6, hal. 4329–4346, 2021. <https://doi.org/10.1109/TAES.2021.3096873>
- [7] J. Schneider, C. Gröger, A. Lutsch, H. Schwarz, dan B. Mitschang, "The lakehouse: State of the art on concepts and technologies," *SN Computer Science*, vol. 5, hal. 449, 2024. <https://doi.org/10.1007/s42979-024-02737-0>
- [8] A. AbouZaid, P. J. Barclay, C. Chrysoulas, dan N. Pitropakis, "Building a modern data platform based on the data lakehouse architecture and cloud-native ecosystem," *Discover Applied Sciences*, vol. 7, hal. 166, 2025. <https://doi.org/10.1007/s42452-025-06545-w>
- [9] C. Yang, C.-H. Wu, J.-C. Shao, Y. Wang, dan C.-M. Hsieh, "AIS-based intelligent vessel trajectory prediction using Bi-LSTM," *IEEE Access*, vol. 10, hal. 24302–24315, 2022. <https://doi.org/10.1109/ACCESS.2022.3154812>
- [10] I W. Sudiarsa, I M. O. Widyantara, N. M. A. E. D. Wirastuti, dan M. Sudarma, "A comparative analysis of LSTM architectures for ship trajectory prediction using AIS data," dalam *Proc. 2025 Int. Conf. Smart-Green Technology in Electrical and Information Systems (ICSGTEIS)*, Bali, Indonesia, Okt. 2025, hal. 118–121. <https://doi.org/10.1109/ICSGTEIS68532.2025.11284502>
- [11] I W. Sudiarsa, I M. O. Widyantara, M. Sudarma, dan N. M. A. E. D. Wirastuti, "Optimizing vessel routing and sustainability with lakehouse architectures: A systematic review," dalam *Int. Proc. Conf. Information Technology, Multimedia, Architecture, Design, and E-Business (IMADE)*, vol. 3, Bali, Indonesia, Des. 2024, hal. 565–574.
- [12] S. Park, C.-S. Yang, dan J. Kim, "Design of vessel data lakehouse with big data and AI analysis technology for vessel monitoring system," *Electronics*, vol. 12, no. 8, hal. 1943, 2023. <https://doi.org/10.3390/electronics12081943>
- [13] M. Cherradi dan A. El Haddadi, "Data lakehouse: Next generation information system," *Seminars in Medical Writing and Education*, vol. 3, hal. 67, 2024. <https://doi.org/10.56294/mw202467>