

Penggunaan Data AIS untuk Monitoring Aktivitas Perkapalan di Perairan Indonesia serta Kunjungan ke Pelabuhan Luar Negeri

222011349, Ladisa Busaina

Dosen Pembimbing: Prof. Setia Pramana, S.Si., M.Sc., Ph.D.

Ringkasan—Badan Pusat Statistik (BPS) memanfaatkan *big data* untuk meningkatkan akurasi dan kecepatan data statistik. Salah satu bentuk *big data* yang populer adalah *Automatic Identification System (AIS)*, yang digunakan untuk memantau dan mengomunikasikan pergerakan kapal. Penelitian ini bertujuan membangun algoritma untuk memonitor aktivitas perkapalan di perairan Indonesia, termasuk kapal yang masuk dan keluar, durasi di pelabuhan, dan kunjungan ke luar negeri. Implementasinya menggunakan pendekatan *Area of Interest (AOI)* berbasis *Distance-Based* dan *Cluster-Based*, serta *Apache Spark* untuk pengolahan data besar. Penelitian ini juga mengembangkan *Standar Operasional Prosedur (SOP)* untuk *preprocessing data AIS*, mencakup validasi, deteksi dan penanganan data hilang atau tidak konsisten, serta koreksi kesalahan manual. Salah satu langkahnya adalah memastikan format *Maritime Mobile Service Identity (MMSI)* benar. Dengan demikian, akurasi dan keandalan data AIS sebagai statistik resmi meningkat, memberikan nilai tambah dalam operasional, prediksi pasar, dan analisis dampak ekonomis sektor maritim di Indonesia.

Kata Kunci—AIS, Kualitas Data, *Distance-Based AOI*, *Cluster-Based AOI*

I. LATAR BELAKANG

Badan Pusat Statistik (BPS) memiliki peran penting dalam menyediakan data strategis dan berkualitas untuk pemerintah, dunia usaha, dan masyarakat. BPS bertanggung jawab untuk menyajikan data yang akurat, tepat waktu, dan relevan guna mendukung program pemerintah dan bisnis dalam persaingan global. Untuk mencapai hal ini, diperlukan sumber daya manusia yang kompeten dan teknologi yang memadai, serta inovasi dalam pencarian dan pengolahan data [1] [2] [3].

Dengan perkembangan teknologi digital, BPS mulai memanfaatkan *big data* untuk meningkatkan akurasi dan kecepatan penyajian data statistik. Salah satu *big data* yang digunakan adalah *Automatic Identification System (AIS)*, suatu sistem otomatis untuk komunikasi dan navigasi kapal. Data AIS memungkinkan pemantauan yang akurat terhadap pergerakan kapal yang berperan penting dalam keselamatan dan pengaturan lalu lintas di laut [4] [5] [6].

Penggunaan data AIS telah terbukti efektif dalam menghasilkan statistik baru, seperti pemantauan lalu lintas kapal, estimasi volume perdagangan, dan analisis dampak ekonomi sektor maritim [7]. Ini meningkatkan akurasi dan keandalan statistik dengan mengurangi kesalahan manusia dalam pengumpulan data [8]. Pemanfaatan AIS di Indonesia diatur oleh Peraturan Menteri Perhubungan Nomor 7 Tahun 2019, yang mewajibkan semua kapal di perairan Indonesia

untuk menggunakan perangkat AIS dan memberikan informasi yang benar [7] [9] [10] [11].

Sebagai negara kepulauan dengan garis pantai panjang, Indonesia sangat diuntungkan dengan penggunaan AIS [12]. AIS membantu memantau aktivitas ilegal seperti penyelundupan dan penangkapan ikan ilegal [11]. Namun, terdapat tantangan dalam melacak kapal kecil di bawah 24 meter dan menghadapi pemalsuan dokumen kapal [13] [14]. Ini menunjukkan perlunya penjaminan kualitas data AIS sebelum digunakan sebagai statistik resmi [8].

Penelitian ini bertujuan untuk membangun algoritma yang memanfaatkan data AIS untuk memantau aktivitas perkapalan di perairan Indonesia. Algoritma ini akan memantau kapal yang masuk dan keluar dari Indonesia, durasi kapal di pelabuhan, dan kunjungan kapal Indonesia ke luar negeri. Implementasi pembangunan algoritma-algoritma tersebut menggunakan pendekatan *Area of Interest (AOI)* berbasis *Distance-Based* dan *Cluster-Based* serta *Apache Spark* untuk pengolahan data besar [7].

Untuk memastikan kualitas data, penelitian ini juga akan mengembangkan *Standar Operasional Prosedur (SOP)* tahap *preprocessing data AIS*. SOP ini mencakup validasi data, deteksi dan penanganan data hilang atau tidak konsisten, serta koreksi kesalahan manual. Salah satu langkahnya adalah memastikan bahwa format *Maritime Mobile Service Identity (MMSI)* yang diinputkan. Proses *preprocessing* yang terstruktur ini diharapkan dapat meningkatkan akurasi dan keandalan data AIS sebagai statistik resmi, memberikan nilai tambah dalam operasional, prediksi pasar, dan analisis dampak ekonomi sektor maritim di Indonesia.

II. TUJUAN PENELITIAN

Penelitian ini memiliki tujuan untuk:

1. Membuat *Standar Operasional Prosedur (SOP)* tahap *preprocessing* sebagai bentuk penjaminan kualitas (*Quality Assurance*) pada data AIS.
2. Membangun algoritma berdasarkan seluruh pelabuhan yang ada di Indonesia untuk:
 - a. mengetahui seluruh kapal yang masuk ke Indonesia; dan
 - b. mengetahui kapal asing yang keluar dari Indonesia.
3. Membangun algoritma untuk mengetahui waktu yang dihabiskan suatu kapal saat berada di pelabuhan Indonesia.
4. Membangun algoritma untuk mengetahui daerah (luar negeri) yang dikunjungi kapal Indonesia.

III. PENELITIAN TERKAIT

Badan Pusat Statistik memaparkan cara mendefinisikan beberapa hal yang berkaitan dengan pengolahan data AIS, di antaranya pendefinisian pelabuhan, pendefinisian jenis kapal, dan pendefinisian rute. Untuk pendefinisian pelabuhan, digunakan data titik pelabuhan dari WPI dan *bounding box* sekitar 12 mil dari titik yang disediakan dalam WPI. Untuk melihat jenis kapal yang dalam *dataset* yang diakses melalui Platform Global PBB (UNGP), pengguna dapat mengakses kolom *vessel_type*. Terakhir, untuk pendefinisian rute, dilakukan identifikasi terhadap aktivitas pelabuhan. Metode yang dilakukan antara lain *geofencing* dengan membuat *geofences* di sekitar pelabuhan, analisis kecepatan dan arah, aktivitas berlabuh, jenis kapal tertentu, dan jenis pesan AIS, seperti laporan posisi dan laporan status navigasi [8].

Asian Development Bank menemukan potensi dari indikator yang diturunkan dari AIS sebagai alternatif untuk *official statistics* maritim yang dapat memberikan informasi penting sebelum *official statistics* maritim dirilis secara resmi. *Events of Interest* (EOI) dan *Areas of Interest* (AOI) digunakan sebagai komponen dasar dari indikator turunan AIS ini. Metode-metode ini mampu mengatasi tantangan umum dalam pemanfaatan data AIS, seperti kualitas data, pengolahan data besar, dan identifikasi batas-batas geografis. Penelitian ini mengidentifikasi kekurangan yang ditimbulkan oleh data AIS, seperti potensi kesenjangan sinyal dari transponder yang dimatikan, variasi dalam pengambilan sampel metode pengambilan sampel, kesalahan dalam bidang input secara manual, MMSI yang salah atau pengidentifikasi kapal, dan kemungkinan pemalsuan atau manipulasi. Penelitian ini juga menemukan bahwa untuk Pelabuhan Nuku'alofa, data AIS secara akurat mencerminkan lonjakan jumlah kapal dari Januari hingga Februari 2022, termasuk kapal angkatan laut dan kapal penyelamat yang memberikan bantuan setelah letusan gunung berapi Hunga Tonga-Hunga Ha'apai di Tonga [7].

Bonham et al. juga melakukan penelitian mengenai pemanfaatan data AIS. Penelitian ini dilakukan dengan pendekatan *machine learning* untuk memprediksi kemungkinan kapal akan tertunda tiba di pelabuhan di Inggris. Selain itu, dilakukan juga eksplorasi terhadap pengoperasian, pemanfaatan, dan hubungan antar pelabuhan di Inggris pada tingkat makro dan perilaku serta karakteristik operasional kapal di level mikro. XGBoost terbukti memberikan kinerja sedikit lebih baik dan digunakan sebagai algoritma pembuktian konsep yang ada. Pendekatan dengan XGBoost juga terbukti lebih bermanfaat secara akurat menangkap data nonlinier dan dapat meningkatkan kinerja model. Model ini dapat digunakan dalam pemuatan pelabuhan simulasi, perencanaan skenario, dan optimalisasi operasi pelabuhan [15].

Terakhir, Alex Noyvirt et al. melakukan eksplorasi penggunaan pergerakan kapal di UK untuk membentuk suatu indikator. Indikator-indikator ini memiliki kemungkinan untuk melengkapi indikator-indikator mengenai aktivitas perdagangan internasional. Indikator-indikator tersebut adalah *'time-in-port'* yang didasarkan pada waktu agregat yang dihabiskan oleh kapal di 10 pelabuhan utama di Inggris dan

'total traffic' yang didasarkan pada jumlah kapal unik yang masuk 10 pelabuhan utama di Inggris [16].

Indikator *'time-in-port'* dihitung dengan menjumlahkan semua periode waktu yang dihabiskan kapal yang memiliki status di dalam pelabuhan sesuai dengan pelabuhannya di setiap pelabuhan dan setiap bulan. Ketika transponder AIS kapal dimatikan di dalam pelabuhan, waktu kontribusinya terhadap indikator hanya dipertimbangkan jika pesan berikutnya yang diterima dari kapal juga berada di dalam pelabuhan yang sama. Aturan ini menghilangkan pencilaan dalam data yang dihasilkan dari kapal yang ditambatkan yang mematikan peralatan AIS mereka dan kemudian meninggalkan pelabuhan tanpa mengaktifkannya kembali atau dari kapal yang karena alasan tertentu mengubah MMSI mereka saat berada di pelabuhan. Kemudian, untuk indikator *'total traffic'*, penghitungannya adalah dengan mengelompokkan data berdasarkan bulan dan status di pelabuhan dan menghitung jumlah kapal yang unik, yang diidentifikasi dengan nomor MMSI mereka. Ketika ada dua atau lebih posisi yang dilaporkan berbeda dalam waktu yang sangat singkat, kemungkinan karena gangguan, jika salah satu dari posisi ini dalam area pelabuhan yang ditentukan, keberadaan kapal dihitung dalam indikator [16].

Indikator-indikator tersebut dianggap penting dalam melengkapi pemahaman mengenai aktivitas perdagangan internasional. Data AIS pada Juli 2016 hingga Agustus 2018 dibandingkan dengan data *official statistics* nilai tambah bruto (NTB) dan statistik perdagangan. Penelitian oleh Alex et al. menemukan korelasi yang sangat baik antara indikator *shipping* dan impor dengan sifat *noisy* dari variabel-variabelnya. Korelasi ini mungkin sebagian besar dipengaruhi oleh variasi musiman. Selain itu, meskipun korelasi keseluruhannya cukup baik, titik-titik yang ada dapat menyimpang jauh. Karena alasan-alasan ini, kami tidak indikator-indikator yang dihasilkan dari penelitian ini tidak direkomendasikan untuk digunakan sebagai prediktor PDB atau statistik ekonomi utama lainnya [16].

IV. METODE PENELITIAN

A. Metode Pengumpulan Data

Pengumpulan data AIS memanfaatkan fasilitas yang disediakan oleh penyedia data AIS exactEarth via Platform Global UN yang dapat diakses pada <https://id.officialstatistics.org/>. Platform ini dapat diakses setelah mendapatkan izin atau hak akses dari pihak UN. Pengajuan hak akses dapat dilakukan dengan mengirimkan surat elektronik menggunakan akun resmi instansi kepada pihak UN, yaitu ke support@officialstatistics.org. Jika pengajuan diterima, pihak UN akan mengirimkan surat elektronik balasan yang berisi izin akses yang diberikan dan *password* yang digunakan untuk mengakses Platform.

Platform Global UN menyediakan *cloud environment* karena menggunakan Amazon Web Service (AWS) sehingga data dapat diakses secara daring melalui Jupyter Hub. Data AIS yang disediakan oleh Platform Global UN terdiri dari data statis, data dinamis, dan data terkait pelayaran yang dikirim kapal dengan perangkat AIS setiap beberapa detik atau menit, tergantung jenis datanya. Data ini berbentuk satu tabel

terstruktur berukuran besar dengan jumlah fitur sebanyak 56. Berikut adalah contoh dari *record* data AIS.

```

-RECORD 1-----
mmsi      | 273295870 | dt_static_utc | 2022-11-19 12:37:44
imo       | 9167758   | dt_insert_utc | 2022-11-19 17:33:12
vessel_name | ARKA-33  | vessel_type_main | null
callsign   | UBPW7     | vessel_type_sub | null
vessel_type | Fishing   | message_type    | 1
vessel_type_cargo | null | source_filename | 53a://ungp-a15-da...
vessel_class | A         | H3_index_0      | 8801ffffffffff
length     | 0.0       | H3_index_1      | 576495936675512319
width      | 0.0       | H3_index_2      | 580968749977385087
flag_country | Russian Federation | H3_index_3      | 585469608825606143
flag_code  | 273       | H3_index_4      | 58999862659368959
destination | BARENTS SEA | H3_index_5      | 594476585331982335
eta        | 8112200   | H3_index_6      | 598988183885611007
draught    | 5.0       | H3_index_7      | 603483783378763775
longitude  | 49.26321667 | H3_index_8      | 607987382905478975
latitude   | 76.78918333 | H3_index_9      | 612498982518161407
sog        | 0.0       | H3_index_10     | 616994582144483327
cog        | 63.3      | H3_index_11     | 621498181771821055
rot        | 352.0     | H3_index_12     | 626801781399183359
heading    | 0.0       | H3_index_13     | 630505381026551295
nav_status | Under Way Using E... | H3_index_14     | 635008980653921599
source     | 0         | H3_index_15     | 639512580281292055
dt_pos_utc | S-AIS     | month           | 11
            |           | day             | 19
  
```

Gambar 1. Contoh Record Data AIS

Platform Global UN juga menyediakan akses terhadap data register kapal dari IHS Markit. Data register kapal ini juga akan digunakan pada penelitian ini. Data utama yang disediakan oleh IHS Markit adalah Nomor Kapal IMO. Selain Nomor Kapal IMO, IHS Markit juga menyediakan informasi rinci mengenai kapal lainnya.

Selain dua data tersebut, pengolahan data AIS juga melibatkan koordinat pelabuhan Indonesia yang kemudian dikembangkan menjadi *Area of Interest* (AOI). Daftar koordinat pelabuhan Indonesia tersebut didapat dengan menerapkan *web scraping* pada situs web *Maritime Safety Information* (MSI). Adapun untuk data statistik resmi yang digunakan adalah data jumlah kunjungan kapal ke Indonesia yang bersumber dari BPS.

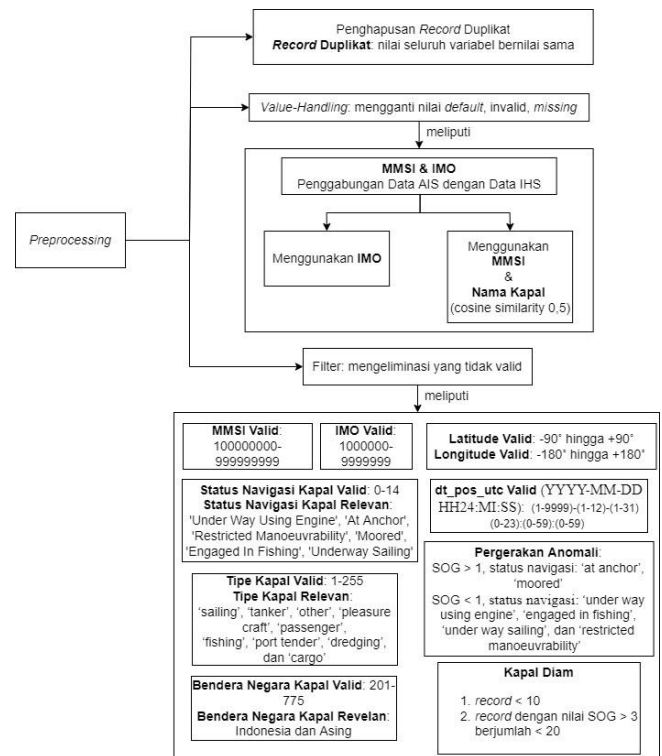
B. Metode Pengolahan Data

Tahap Preprocessing Data

Seperti data lainnya, data AIS juga memiliki permasalahan akan kualitas datanya. Ini memungkinkan terdapat pesan yang memiliki *default value*, *invalid value*, *missing value*, *noise* atau *outlier*. Beberapa hal yang menyebabkan hal tersebut adalah hilangnya sinyal perangkat AIS, *static message* yang rentan terhadap *error*, dan MMSI tidak sesuai dengan ketentuan [7]. Untuk itu, data AIS perlu melalui tahap *preprocessing* yang tepat untuk mengurangi *default value*, *invalid value*, *missing value*, dan *noise* atau *outlier* pada data AIS serta mendapatkan data yang sesuai dengan cakupan penelitian.

Tahap *preprocessing* ini dilakukan dengan menerapkan *value-handling* pada data AIS. *Default value*, *invalid value* atau *noise* atau *outlier*, dan *missing value* yang tidak dapat teratasi dengan *value-handling* akan ditangani menggunakan proses filter sesuai dengan format masing-masing fitur [17] [18] [19] [20]. Filter juga diterapkan agar didapatkan data yang relevan dengan penelitian yang dilakukan. Sebelum *value-handling* dan filter diterapkan, data telah dipastikan merupakan *record* unik dengan melakukan penghapusan *record* duplikat. Hal ini dikarenakan Data AIS sering kali mengandung *noise* berupa data yang duplikat [21]. Berikut skema tahap *preprocessing* yang dilakukan.

Berikut skema tahap *preprocessing* yang dilakukan.



Gambar 2. Skema Tahapan Preprocessing

Pembentukan Algoritma

Dari data AIS yang telah dikumpulkan, akan dibentuk algoritma. Untuk algoritma pengawasan di kawasan pelabuhan, diperlukan definisi yang jelas mengenai kawasan pelabuhan, yang merupakan bagian dari penentuan *Area of Interest* (AOI). Terdapat dua pendekatan yang dapat digunakan, yaitu *Distance-Based Approach* dan *Cluster-Based Approach* [7].

Setelah menentukan AOI, ditentukan juga *Event of Interest* (EOI). EOI dari penelitian ini adalah seluruh aktivitas keluar-masuk pelabuhan Indonesia dan negara utama tujuan ekspor Indonesia, yang dalam hal ini merupakan AOI, oleh seluruh kapal. Dari AOI dan EOI tersebut, dibentuklah indikator mengenai aktivitas perkapalan Indonesia dan di Indonesia. Berikut penjelasan dari metode pembentukan setiap indikator.

1. Jumlah kapal masuk dan keluar (numVisit)

Jumlah kapal masuk dan keluar dapat dihitung dengan menggunakan koordinat pesan AIS yang dikirimkan kapal. Apabila pesan sebelumnya dikirim di luar AOI (pelabuhan) pesan tersebut dihitung satu kunjungan masuk pada pelabuhan tersebut. Fitur yang digunakan untuk membentuk 'numVisit' adalah 'MMSI', 'dt_pos_utc', 'longitude', 'latitude', dan 'H3_index_8'.

2. Waktu kapal di pelabuhan (timeInPort)

Waktu kapal di pelabuhan merupakan rata-rata waktu yang dihabiskan kapal di suatu pelabuhan (AOI) dalam satuan detik. Indikator ini dihitung dengan menjumlahkan semua periode waktu yang dihabiskan kapal yang memiliki status di dalam pelabuhan sesuai dengan pelabuhannya di setiap pelabuhan dan setiap bulan. Ketika transponder AIS kapal dimatikan di dalam pelabuhan, waktu kontribusinya terhadap indikator hanya dipertimbangkan jika pesan

berikutnya yang diterima dari kapal juga berada di dalam pelabuhan yang sama. Aturan ini menghilangkan pencilaan dalam data yang dihasilkan dari kapal yang ditambatkan yang mematikan peralatan AIS mereka dan kemudian meninggalkan pelabuhan tanpa mengaktifkannya kembali atau dari kapal yang karena alasan tertentu mengubah MMSI mereka saat berada di pelabuhan. Fitur yang digunakan untuk membentuuk ‘timeInPort’ adalah ‘MMSI’, ‘dt_pos_utc’, ‘longitude’, dan ‘latitude’, dan ‘H3_index_8’.

C. Metode Analisis Data

Visualisasi Data

Hasil pengaplikasian algoritma yang telah terbentuk pada data AIS akan divisualisasikan. Visualisasi akan dilakukan menggunakan Kepler.gl. Kepler.gl adalah aplikasi berbasis web yang sangat efektif untuk eksplorasi visual data geolokasi dalam skala besar. Dengan menampilkan data geospasial dalam antarmuka web tunggal, Kepler.gl membantu pengguna dengan cepat memvalidasi ide dan mendapatkan wawasan dari visualisasi tersebut [22].

Analisis Deskriptif

Pada penelitian ini, hasil pengolahan data AIS akan dianalisis secara deskriptif. Analisis deskriptif terutama digunakan pada saat melakukan eksplorasi data AIS. Analisis deskriptif memiliki tujuan untuk mengetahui gambaran atau kecenderungan data melalui tabel, grafik, mean, median, modus, atau teknik statistik lain dengan tidak melakukan generalisasi [23].

Evaluasi Performa Algoritma

Salah satu metode yang efektif mempelajari efisiensi algoritma adalah dengan menggunakan Notasi *Big-O*. Notasi *Big-O* didefinisikan sebagai representasi matematis dari pertumbuhan waktu algoritma seiring dengan penambahan ukuran masukan [24].

Notasi *Big-O* memiliki nilai di antaranya $O(1)$, $O(\log(n))$, $O(n)$, $O(n^2)$, dan $O(n^n)$. Berikut adalah penjelasan masing-masing nilai.

1. Notasi *Big-O* akan bernilai $O(1)$ saat algoritma tidak memiliki perulangan dan waktu yang dibutuhkan untuk menjalankan algoritma adalah konstan berapa pun inputnya.
2. Notasi *Big-O* akan bernilai $O(n)$ saat algoritma memiliki satu perulangan untuk setiap baris *item*. Waktu yang dibutuhkan untuk menjalankan algoritma akan meningkat dengan kecepatan yang dengan input.
3. Notasi *Big-O* akan bernilai $O(\log(n))$ selama algoritma rekursif. algoritma memiliki perulangan sebanyak n untuk setiap baris *item*. Waktu yang dibutuhkan untuk waktu berjalan hampir tidak meningkat saat input ditingkatkan secara eksponensial.

Evaluasi Kinerja Algoritma

Pada penelitian ini, algoritma yang telah terbentuk akan diujikan kepada data AIS. Untuk mengetahui keakuratan dari

algoritma, diperlukan evaluasi terhadap hasil dari algoritma tersebut. Teknik untuk mengevaluasi algoritma yang dihasilkan menggunakan RMSE dan MAPE.

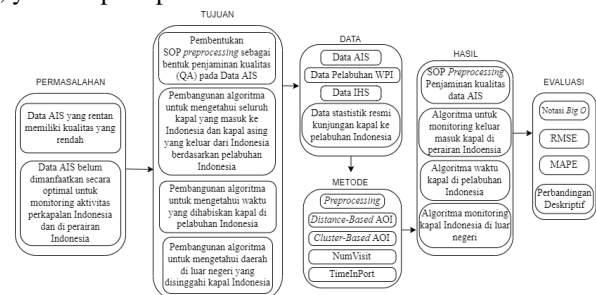
Evaluasi terhadap logaritma juga dilakukan dengan membandingkan data yang dihasilkan oleh algoritma dengan data *official statistics* oleh Badan Pusat Statistik secara deskriptif.

V. KERANGKA PIKIR

Penelitian ini dimulai dengan membentuk suatu Standar Operasional Prosedur (SOP) preprocessing sebagai penjaminan kualitas (QA) pada data AIS yang dihasilkan dari kapal-kapal. Salah satu preprocessing yang dilakukan adalah menggabungkan data AIS dan data IHS yang merupakan data registrasi kapal. Data AIS akan dimanfaatkan dalam penelitian ini akan untuk mengetahui seluruh kapal yang masuk ke perairan Indonesia dan kapal asing yang keluar dari perairan Indonesia berdasarkan pelabuhan Indonesia. Penelitian ini juga dilakukan untuk mengetahui waktu yang dihabiskan di pelabuhan Indonesia serta mengetahui daerah di luar negeri yang disinggahi atau dikunjungi kapal Indonesia. Upaya mencapai hal-hal tersebut adalah dengan membangun algoritma yang sesuai menggunakan pendekatan *Distance-Based AOI* dan *Cluster-Based AOI*.

Algoritma yang dibangun akan diterapkan pada data AIS. Untuk mengetahui tingkat keefesienan algoritma yang dibangun, akan dilakukan evaluasi terhadap performa algoritma dengan Notasi *Big O*. Evaluasi kemudian dilanjutkan untuk mengukur kinerja algoritma melalui beberapa ukuran evaluasi, yaitu RMSE dan MAPE. Kinerja dari algoritma juga akan diukur dengan membandingkan hasil yang didapat oleh algoritma dengan data statistik resmi kunjungan kapal ke pelabuhan Indonesia secara deskriptif. Pada penelitian ini juga akan ditampilkan beberapa statistik mengenai karakteristik kapal dan lalu lintasnya.

Berikut kerangka pikir yang menjadi acuan dalam penelitian ini, yakni seperti pada Gambar 3.



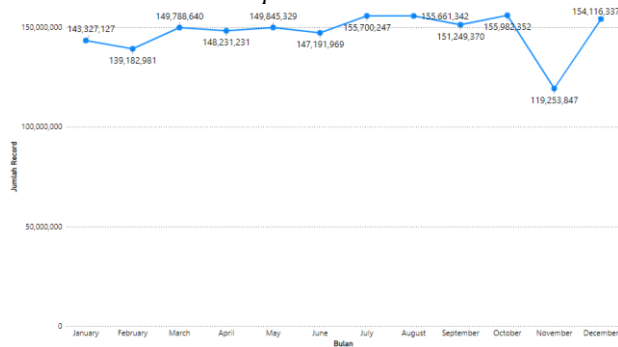
Gambar 3. Kerangka Pikir Penelitian

VI. HASIL DAN PEMBAHASAN

A. Gambaran Umum Data AIS

Eksplorasi dilakukan pada data AIS yang sempat terdeteksi di Indonesia sepanjang tahun 2022. *Record* data AIS tersebut berjumlah 1.769.530.772 *record*.

Jumlah Records Data AIS per Bulan Tahun 2022



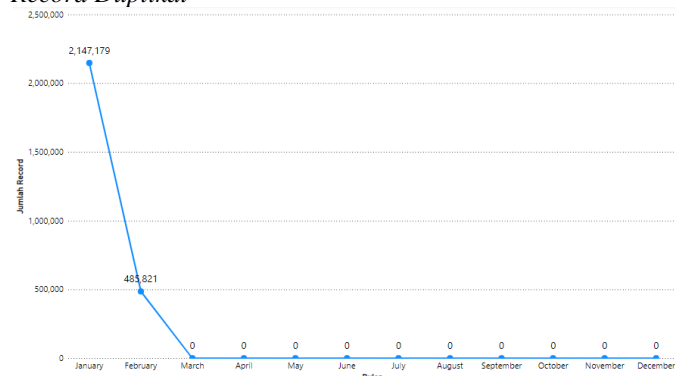
Gambar 4. Jumlah Record Data AIS per Bulan Tahun 2022

Berdasarkan grafik di atas, terlihat bahwa jumlah *records* data AIS yang sempat terdeteksi di Indonesia setiap bulannya selama 2022 cenderung berfluktuatif. Artinya, terdapat jumlah *records* yang sangat tinggi di bulan tertentu dan terdapat jumlah *records* yang sangat rendah di beberapa bulan lainnya.

B. Standar Operasional Prosedur (SOP) Preprocessing sebagai Penjaminan Kualitas Data (Quality Assurance)

Data AIS memerlukan kontrol kualitas (*quality control*) terhadap datanya. Hal ini dilakukan untuk menghindari diolahnya data yang tidak sesuai ketentuan, seperti data dengan *default value*, *invalid value*, *missing value*, dan *noise* atau *outlier*, serta mendapatkan data yang sesuai dengan cakupan penelitian. Berikut Standar Operasional Prosedur (SOP) *preprocessing* yang diterapkan terhadap data AIS dalam bentuk beberapa indikator *quality control*.

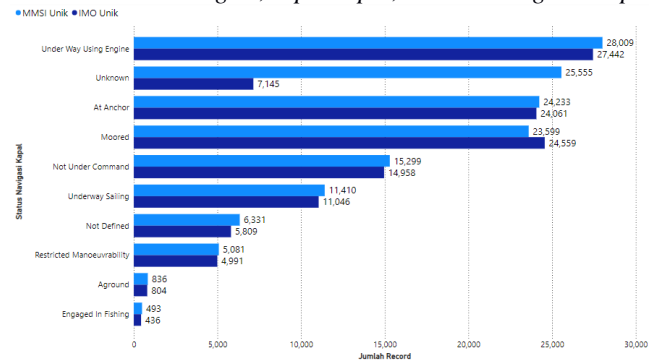
Record Duplikat



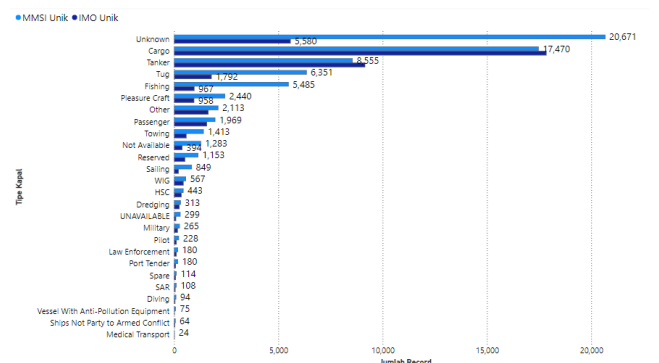
Gambar 5. Jumlah Record Duplikat

Berdasarkan grafik pada Gambar 5, *records* terdeteksi pada awal tahun, yaitu pada bulan Januari dan Februari. Pada bulan Februari, jumlah *records* yang terdeteksi adalah 485.821 yang mengindikasikan adanya penurunan sebesar 77% dibandingkan bulan sebelumnya. Artinya, pada bulan Januari dan Februari, terdapat kapal yang mengirimkan pesan yang sama dan oleh penyedia data ExactEarth tidak dieliminasi.

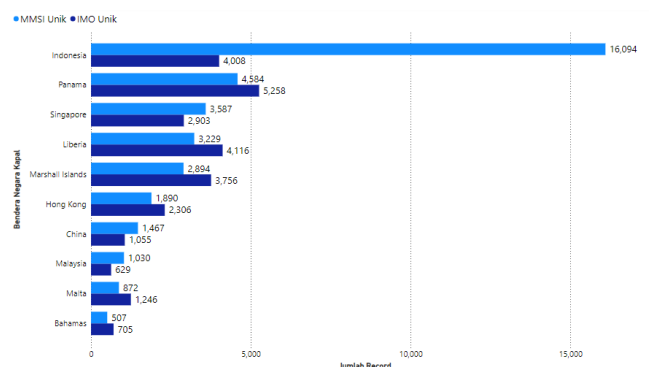
Jumlah Record Data AIS dengan MMSI dan IMO Unik Menurut Status Navigasi, Tipe Kapal, Bendera Negara Kapal



Gambar 6. Jumlah Record Data AIS dengan MMSI dan IMO Unik Menurut Status Navigasi Tahun 2022



Gambar 7. Jumlah Record Data AIS dengan MMSI dan IMO Unik Menurut Tipe Kapal Tahun 2022



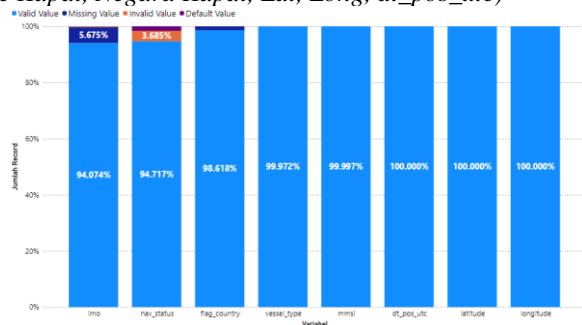
Gambar 8. Jumlah Record Data AIS dengan MMSI dan IMO Unik Menurut Bendera Negara Kapal Tahun 2022

Berdasarkan diagram pada Gambar 6, 7, dan 8 tersebut, terdapat beberapa dari status navigasi, tipe, dan bendera negara kapal yang memiliki jumlah *records* dengan IMO unik kurang dari jumlah *records* dengan MMSI unik. Hal ini dapat terjadi karena adanya kapal yang memiliki perangkat AIS sehingga memiliki nomor MMSI, tetapi tidak memiliki nomor IMO sebagai identitas kapal.

Selain untuk identifikasi, nomor IMO adalah alat utama bagi berbagai aktivitas industri transportasi laut. Menurut peraturan SOLAS (Safety of Life at Sea), memiliki nomor IMO

adalah prasyarat untuk berlayar dan menjual [25]. Jadi, ada beberapa kemungkinan penyebab terjadinya ketimpangan jumlah *records* dengan IMO unik dan jumlah *records* dengan MMSI unik dilihat dari tipe kapalnya, yaitu tipe kapal tersebut tidak diwajibkan untuk memiliki nomor IMO, kapal dengan tipe kapal yang diwajibkan untuk memiliki nomor IMO memiliki tonase kotor (GT) kurang dari 100 ton, dan kapal dengan tipe kapal dan berat yang diwajibkan memiliki nomor IMO tidak memiliki nomor IMO tersebut.

Proporsi Nilai Valid, Default, Invalid, dan Missing Record Data AIS Menurut Variabel (MMSI, IMO, Status Navigasi, Tipe Kapal, Negara Kapal, Lat, Long, dt_pos_utc)



Gambar 9. Proporsi Nilai Valid, Default, Invalid, dan Missing Record Menurut Variabel

Berdasarkan diagram pada Gambar 9 tersebut, terdapat tiga variabel atau fitur dengan nilai valid kurang dari 99%, yaitu IMO, status navigasi, dan bendera kapal. Sebesar 5,68% dari keseluruhan nilai IMO merupakan missing value, sedangkan status navigasi memiliki nilai invalid sebesar 3,69% dari keseluruhan datanya. Bendera negara juga memiliki beberapa nilai yang merupakan missing value, yaitu sekitar 1,38% dari keseluruhan datanya. Ini artinya perlu penanganan lebih lanjut terhadap nilai missing, default, dan invalid fitur IMO.

Pergerakan Anomali



Gambar 10. Jumlah Record yang Diindikasikan sebagai Pergerakan Anomali

Berdasarkan grafik pada Gambar 10, *records* yang diindikasikan sebagai pergerakan anomali memiliki pola fluktuatif. Artinya, terdapat bulan tertentu dengan jumlah *records* yang terindikasi sebagai pergerakan anomali yang sangat tinggi dan terdapat beberapa bulan lainnya dengan jumlah *records* yang terindikasi sebagai pergerakan anomali yang sangat rendah.

Secara keseluruhan, hasil dari penjaminan kualitas untuk data AIS adalah sebagai berikut.

TABEL I
TABEL HASIL PENJAMINAN KUALITAS DATA AIS

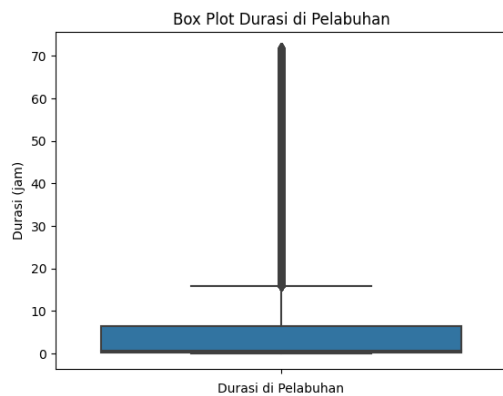
Penjaminan Kualitas	Baik	Kurang Baik	Keterangan
Records Duplikat		✓	Terdapat <i>records</i> yang merupakan <i>record</i> duplikat
Jumlah <i>records</i> dengan MMSI dan IMO Unik menurut Status Navigasi		✓	Jumlah <i>records</i> dengan MMSI dan IMO unik belum seimbang
Jumlah <i>records</i> dengan MMSI dan IMO Unik menurut Tipe Kapal		✓	Jumlah <i>records</i> dengan MMSI dan IMO unik belum seimbang
Jumlah <i>records</i> dengan MMSI dan IMO Unik menurut Negara Kapal		✓	Jumlah <i>records</i> dengan MMSI dan IMO unik belum seimbang
Jumlah <i>records</i> dengan nilai default, invalid, missing menurut MMSI, IMO, nav_status, vessel_type, flag_country, latitude, longitude, dt_pos_utc		✓	Terdapat fitur dengan nilai <i>default</i> , <i>invalid</i> , atau <i>missing value</i>
Pergerakan anomali		✓	Banyaknya jumlah pergerakan anomali
Kapal diam		✓	Terdapat kapal (MMSI) yang diindikasikan sebagai kapal diam

Dapat disimpulkan bahwa kualitas data AIS masih belum baik. Karenanya, perlu dilakukan *preprocessing* sebagai bentuk dari penjaminan kualitas data AIS.

Pertama, dilakukan penghapusan *record* duplikat. Tahap ini menyisakan sebanyak 1.766.897.772 *record*. Artinya, sebanyak 0,15% *record* telah berkurang dibandingkan jumlah *record* AIS yang sempat terdeteksi di Indonesia sepanjang tahun 2022 yang berjumlah 1.769.530.772 *record*. Pada tahap ini, masih terdapat fitur-fitur yang memiliki nilai *default*, *invalid*, *missing*.

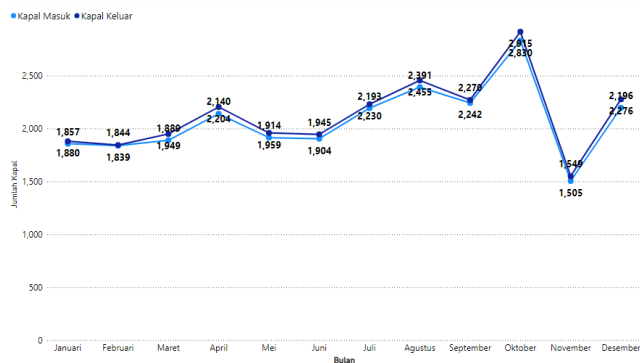
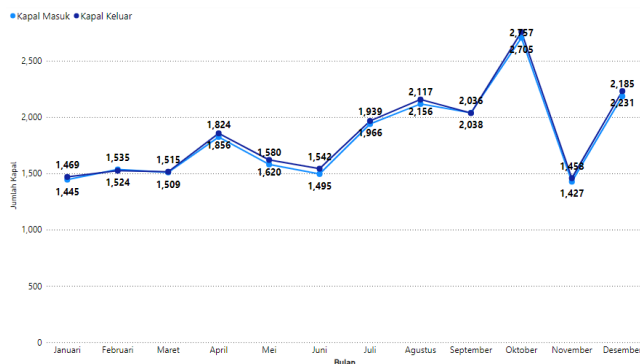
Untuk menangani hal tersebut, dilakukan penggabungan data AIS dengan data IHS sebagai bentuk *value-handling* dari fitur IMO dan MMSI. Penggabungan data dilakukan dengan mencocokkan IMO serta MMSI dan nama kapal dengan kemiripan di atas 0,5 yang diukur menggunakan *cosine similarity*. *Record* hasil penggabungan data AIS dan data IHS berjumlah 1.696.875.705 *record*. Artinya, sebanyak 4,11% *record* telah berkurang dibandingkan jumlah *record* AIS hasil penghapusan *record* duplikat yang berjumlah 1.766.897.772 *record*.

Kemudian, beberapa langkah filter diterapkan terhadap data AIS untuk mengeliminasi *default value*, *invalid value*, *missing value*, dan *noise* atau *outlier*, serta mendapatkan data yang terkait penelitian ini, seperti yang sudah dijelaskan pada bab IV. TABEL II menunjukkan jumlah pesan AIS pada setiap tahap filter, mulai dari sebelum difilter hingga jumlah data setelah difilter.

Gambar 17 Distribusi Durasi Kapal di Pelabuhan Indonesia (*Cluster-Based AOI*)

Kedua pendekatan menghasilkan durasi kapal di pelabuhan Indonesia dengan distribusi yang sama sepanjang tahun 2022, yaitu berada pada rentang 0 hingga 20 jam, dan memiliki beberapa durasi ekstrem, yaitu antara 20-72 jam.

Algoritma Deteksi Kunjungan Indonesia ke Luar Negeri

Gambar 18 Jumlah Kunjungan Kapal Indonesia ke Luar Negeri per Bulan (*Distance-Based AOI*)Gambar 19 Jumlah Kunjungan Kapal Indonesia ke Luar Negeri per Bulan (*Cluster-Based AOI*)

Kedua pendekatan menghasilkan jumlah kunjungan kapal Indonesia ke luar negeri dengan pola yang sama sepanjang tahun 2022, yaitu tertinggi pada bulan Oktober dan terendah pada bulan November.

H. Evaluasi Performa Algoritma

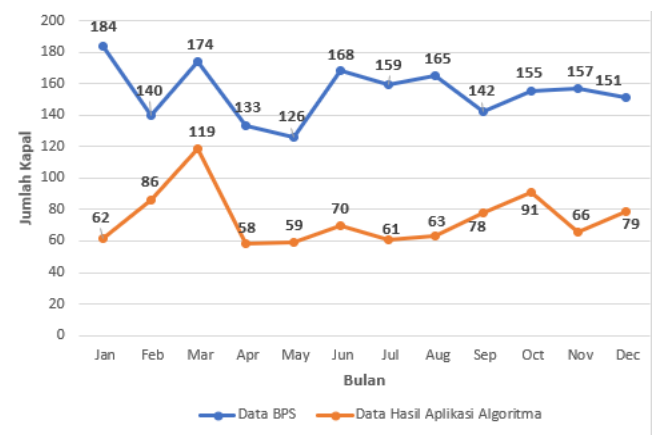
Berdasarkan hasil implementasi, dilakukan evaluasi performa algoritma dengan menghitung RMSE dan MAPE. Evaluasi ini hanya dilakukan pada Algoritma Arus Masuk-Keluar Kapal di Pelabuhan Indonesia karena keterbatasan pada sumber data statistik resmi yang dimiliki. Berikut ditampilkan hasil penghitungannya.

TABEL III
TABEL EVALUASI KINERJA ALGORITMA PADA BEBERAPA PELABUHAN INDONESIA

Pelabuhan yang Dikunjungi	<i>Distance-Based AOI</i>		<i>Cluster-Based AOI</i>	
	RMSE	MAPE	RMSE	MAPE
(1)	(2)	(3)	(4)	(5)
Teluk Bayur	83,31866538	52,07228907	82,76875417	51,65349462
Cirebon	148,1350735	91,21597754	142,3294301	87,80385546
Banten	1603,965658	133,6675182	4895,368151	412,9450644
Benoa	70,11003257	47,03633347	70,44527306	53,2503679
Pontianak	191,1661581	74,58477038	217,1898171	89,71506381
Poso	11,41271221	70,30092593	13,40708768	85,86805556
Amamapare	83,7744989	400,9215371	127,3908029	612,0061837
70 Pelabuhan	8624,301405	42,27629581	8963,522752	43,93919217

Berdasarkan hasil penghitungan RMSE dan MAPE tersebut, diketahui bahwa algoritma menggunakan *Distance-Based AOI* menghitung jumlah kapal yang berkunjung ke pelabuhan Indonesia lebih baik daripada hasil penghitungan menggunakan algoritma *Cluster-Based AOI*.

Selain menghitung RMSE dan MAPE pada jumlah kunjungan kapal ke beberapa pelabuhan di Indonesia, ditampilkan juga perbandingan jumlah kunjungan kapal di pelabuhan Teluk Bayur menggunakan diagram garis, seperti yang terlihat pada grafik Gambar 20. Berdasarkan grafik tersebut, dapat dilihat masih adalah perbedaan jumlah kunjungan kapal di pelabuhan Teluk Bayur yang telah dihitung menggunakan algoritma *Cluster-Based AOI* dengan jumlah kunjungan kapal di pelabuhan Teluk Bayur yang telah dihitung oleh Badan Pusat Statistik.

Gambar 20 Perbandingan Jumlah Kunjungan di Pelabuhan Teluk Bayur Menurut BPS dan Hasil Penghitungan Menggunakan Algoritma *Cluster-Based AOI*

VII. PENUTUP

Berdasarkan hasil penelitian dan pembahasan yang telah dipaparkan, berikut beberapa kesimpulan yang dapat diambil dari penelitian ini.

1. Data AIS dapat digunakan untuk membangun algoritma terkait pemantauan aktivitas kapal di lautan. Dengan menerapkan Standar Operasional Prosedur (SOP) *preprocessing* yang tepat, hasil implementasi dari algoritma yang terbentuk dari data AIS dapat mendekati data statistik kunjungan kapal ke pelabuhan Indonesia.
2. Beberapa algoritma yang terbentuk dari data AIS, yaitu deteksi arus masuk-keluar kapal di pelabuhan Indonesia, penghitungan durasi kapal di pelabuhan Indonesia, dan deteksi kunjungan Indonesia ke luar negeri.
3. Algoritma yang terbentuk dari data AIS dapat digunakan sebagai alternatif data transportasi laut, seperti jumlah kunjungan kapal domestik dan luar negeri ke pelabuhan Indonesia, durasi kapal di pelabuhan Indonesia, dan jumlah kunjungan kapal Indonesia ke pelabuhan luar negeri.
4. Pendekatan *Distance-Based* dan *Cluster-Based* digunakan untuk membantu Area of Interest pelabuhan. Secara umum, *Distance-Based* AOI merupakan pendekatan terbaik untuk menghitung jumlah kunjungan kapal domestik dan luar negeri ke pelabuhan Indonesia. Namun, terdapat beberapa pelabuhan yang menghasilkan statistik yang lebih baik jika menggunakan *Cluster-Based* AOI.

Selain itu, diberikan juga beberapa saran dari penelitian ini untuk penelitian selanjutnya, yaitu sebagai berikut.

1. Menggunakan data AIS dari penyedia data lain selain exactEarth untuk melihat perbedaan pengolahan data sebelum disediakan kepada pengguna data.
2. SOP *preprocessing* data AIS perlu dikaji atau dikembangkan lebih lanjut, seperti mengidentifikasi karakteristik kapal diam lebih detail agar hasil perhitungan yang didapatkan lebih relevan.
3. Menggunakan metode pendekatan *Area of Interest* pelabuhan yang lain, seperti *Manual* AOI dengan pertimbangan data luas pelabuhan diketahui.
4. Data hasil implementasi algoritma yang terbentuk dari data AIS ini bisa digunakan sebagai komplemen statistik transportasi laut oleh Tim Jasa Transportasi Domestik Direktorat Statistik Distribusi dan Direktorat Analisis dan Pengembangan Statistik.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik, "Arah Perubahan Badan Pusat Statistik 2021-2024," *Badan Pus. Stat.*, pp. 1–125, 2021.
- [2] Badan Pusat Statistik. (2024). Mengapa Harus Big Data?. 12-06-2024 <https://bigdata.bps.go.id/about>
- [3] BPS Provinsi Jambi. (2022). Menuju Data yang Berkualitas di Era Big Data. 12-06-2024 <https://jambi.bps.go.id/news/2022/08/02/200/menuju-data-yang-berkualitas-di-era-big-data.html>
- [4] R. Haryadi, H. Setiawan, W. Hermawansyah, M. Masmilah, and S. Bani Saleh, "Sistem Penguraian Data Automatic Identification System (AIS) dengan Bahasa Pemrograman Python," *Semin. Nas. Rekayasa dan Teknol.*, vol. 27, no. November, pp. 18–23, 2019.
- [5] A. Maulidi, "Disain Sistem Navigasi Automatic Identification System (AIS) Transceiver Berbasis Mini Computer Pada Kapal Nelayan Tradisional Di Madura," *Inovtek Polbeng*, vol. 9, no. 1, p. 12, 2019, doi: 10.35314/ip.v9i1.878.
- [6] A. Karim, "(AUTOMATIC IDENTIFICATION SYSTEM (AIS) SATELIT DATA CORRECTION USING INTERPOLATION AND EXTRAPOLATION METHODE, (Case Study: LAPAN-A2 and LAPAN-A3 Satellite)," *J. Teknol. Dirgant.*, vol. 16, no. 2, p. 159, 2019, doi: 10.30536/j.jtd.2018.v16.a3049.
- [7] A. D. Bank, "Methodological Framework For Unlocking Maritime Insights Using Automatic Identification System Data," 2023.
- [8] B. P. Statistik, "Kajian Pemanfaatan Data Automatic Identification System (AIS)," 2023.
- [9] J. Pence, "International maritime organization," *Int. J. Mar. Coast. Law*, vol. 24, no. 4, pp. 713–725, 2009, doi: 10.1163/092735209X12499043518304.
- [10] K. E. Y. Points and K. Kim, "Economic Applicability of the Automatic Identification System Data: Use Cases and the Way Forward," 2023. [Online]. Available: <http://dx.doi.org/10.22617/BRF230082-2>.
- [11] S. Simau, I. Prakoso, J. I. Manengkey, J. Manohas, P. Pontoh, and G. K. da Gomez, "Kajian Pemanfaatan Ais (Automatic Identification System) Dalam Melacak Aktivitas Illegal Fishing Pada Kapal Dan Ais Hybrid Sebagai Alat Bantu Penangkapan Ikan," *J. Bluefin Fish.*, vol. 4, no. 2, pp. 102–128, 2023.
- [12] M. F. Arianto, "Potensi Wilayah Pesisir di Negara Indonesia," *J. Geogr.*, vol. 20, no. 20, pp. 1–7, 2020.
- [13] Indonesia Ocean Justice Initiative. (2021, 2 Juli). IUU Fishing Di Laut Natuna Utara, Laut Sulawesi Dan Selat Malaka – Juni 2021. 10-10-2023 <https://oceanjusticeinitiative.org/2021/07/02/iuu-fishing-di-natuna-sulawesi-selat-malaka-juni-2021/>
- [14] Maulana, A. (2017, 31 Maret). Indonesia Darurat IUU Fishing. Kantor Komunikasi Publik. 10-10-2023 <https://www.unpad.ac.id/2017/03/indonesia-darurat-iuu-fishing/>
- [15] C. Bonham, A. Noyvirt, I. Tsalamanis, and S. Williams, "Analysing port and shipping operations using big data," *ONS Data Sci. Campus*, pp. 1–50, 2018, [Online]. Available: <https://datasciencecampus.ons.gov.uk/wp-content/uploads/sites/10/2018/06/here.pdf>.
- [16] L. N. Alex Noyvirt, Ioannis Kaloskampis, Stephen Campbell, Sumit Dey-Chowdhury, "Faster Indicators of UK Economic Activity: Shipping," *Data Science for The Public Good*, 2019. <https://datasciencecampus.ons.gov.uk/projects/faster-indicators-of-uk-economic-activity-shipping/>.
- [17] International Telecommunication Union, "Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile band (Recommendation ITU-R M.1371-4)," *Int. J. Rock Mech. Min. Sci.*, vol. 42, no. 4, pp. 481–507, 2005, [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.1371-5-201402-I!!PDF-E.pdf.
- [18] D. Salgado, B. Oancea, M. Necula, and J. D. Salgado, "ESSnet Big Data II Specific Grant Agreement No 1 (SGA-1)," 2017. [Online]. Available: <https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata>.
- [19] Raymond, E. S. (2021, Januari). AIVDM/AIVDO Protocol Decoding. 26-01-2024 <https://gpsd.gitlab.io/gpsd/AIVDM.html>
- [20] SINAY Maritime Data Solution. (2022, Juli). What is the Difference Between IMO and MMSI?. 04-03-2024 <https://sinay.ai/en/what-is-the-difference-between-imo-and-mmsi/>
- [21] T. Emmens, C. Amrit, A. Abdi, and M. Ghosh, "The promises and perils of Automatic Identification System data," *Expert Syst. Appl.*, vol. 178, p. 114975, 2021, doi: 10.1016/j.eswa.2021.114975.
- [22] P. R. Koch, C. A. Hölbig, and R. Rieder, "Using Kepler.gl to visualize weather data," 2018.
- [23] I. A. Siregar, "Analisis Dan Interpretasi Data Kuantitatif," *ALACRITY J. Educ.*, vol. 1, no. 2, pp. 39–48, 2021, doi: 10.52121/alacrity.v1i2.25.
- [24] Ichi.Pro. (2024). Notasi Big-O: Penjelasan Sederhana dengan Contoh. 10-03-2024 <https://ichi.pro/id/notasi-big-o-penjelasan-sederhana-dengan-contoh-181845825280714>
- [25] Shine Micro Inc. (2024). MMSI (Maritime Mobile Service Identity) Overview. 04-03-2024 <https://www.shinemicro.com/about-mmsi/>