

Prediction of Skipjack Tuna Catches Using Linear Regression and Recurrent Neural Network Methods

Prediksi Hasil Tangkapan Ikan Cakalang dengan Metode Regresi Linear dan Recurrent Neural Network

Theofilio K. Tarumingkeng¹, Sherwin R.U.A Sompie², Agustinus Jacobus³, Wilhelmina Patty⁴

^{1,2,3} Dept. of Electrical Engineering, Sam Ratulangi University Manado, Kampus Bahu St., 95115, Indonesia

⁴ Faculty of Fisheries and Marine Sciences, Sam Ratulangi University, Manado-95115

e-mails : theotarumingkeng21@gmail.com, aldo@unsrat.ac.id, a.jacobus@unsrat.ac.id,
wilhelmina.patty@unsrat.ac.id

Received: 19 January 2024 ; revised: 25 February 2024 ; accepted: 30 March 2024

Abstract – Skipjack tuna is one of the abundant pelagic fish resources in the Sulawesi Sea. The aim of this research is to predict the production of skipjack tuna catches by applying the Linear Regression and RNN methods. The research was carried out from March 2023 to December 2023. The data used were catches of skipjack that were landed at the Tumumpa Beach Fishing Harbor, North Sulawesi from 2018 to 2023. Data on environmental factors was taken from the Marine Copernicus and Climate Copernicus websites in June 2023, in 14 observation points. The model was created using the Scikit-learn library and the Keras library which is included in the Tensorflow library. The results of the analysis shows that there is a significant relationship between several environmental variables, namely between SST and the Chlorophyll, which is of a negative nature. However, the relationship between the environmental variables analyzed and skipjack tuna catches appears to be relatively low (less than 25%) of the remainder due to the influence of other factors. The modeling results show that the predictor variables used (SST, Chlorophyll, Sea Breeze, and Sea Currents) have a less significant influence (less than 50%) on skipjack tuna catches. The performance of both models is still not satisfactory in predicting skipjack tuna catches.

Keywords – Catches; Linear Regression; Prediction; RNN.

Abstrak – Ikan Cakalang (*Katsuwonus Pelamis*) merupakan salah satu sumber daya ikan pelagis yang melimpah di Laut Sulawesi. Tujuan penelitian ini adalah memprediksi produksi hasil tangkapan ikan Cakalang dengan menerapkan metode Regresi Linear dan Recurrent Neural Network (RNN). Penelitian dilaksanakan dari Maret 2023 sampai Desember 2023. Data yang digunakan adalah hasil tangkapan cakalang yang didaratkan di Pelabuhan Perikanan Pantai Tumumpa, Sulut dari 2018 sampai 2023. Data faktor lingkungan diambil dari situs web *Marine Copernicus* dan *Climate Copernicus* pada bulan Juni 2023, di 14 titik pengamatan. Model dibuat dengan menggunakan library *Scikit-learn* dan library *Keras* yang telah disertakan dalam library *Tensorflow*. Hasil analisis menunjukkan bahwa ada hubungan signifikan diantara beberapa variabel lingkungan yakni antara variabel SPL dan variabel klorofil, yang bersifat negatif. Namun hubungan antara variabel lingkungan yang dianalisis dan hasil tangkapan ikan cakalang, terlihat relatif rendah, yakni hanya dapat dijelaskan kurang dari 25%, selebihnya karena pengaruh faktor lain. Hasil permodelan menunjukan bahwa variabel prediktor yang digunakan (SPL, Klorofil, Angin Laut, dan Arus Laut) memiliki pengaruh yang kurang signifikan (kurang dari 50%) terhadap hasil tangkapan ikan Cakalang. Kinerja kedua model masih belum memuaskan dalam memprediksi hasil tangkapan ikan Cakalang.

Kata Kunci –Hasil tangkapan; Prediksi RNN; Regresi Linear.

I. PENDAHULUAN

Data potensi perikanan tangkap sangat besar baik jumlah dan keragamannya, yang sangat penting sebagai salah satu sumber modal utama pembangunan bila dikelola dengan baik, bertanggung jawab dan berkelanjutan [1]. Untuk itu diperlukan kebijakan penangkapan ikan terukur berbasis kuota yang terkuantifikasi/terukur secara akurat [2].

Data produksi perikanan tangkap di PPI Tumumpa setiap tahun sangat banyak [3], namun pengelolaan data tersebut saat ini dinilai belum optimal untuk pengambilan keputusan pengelolaan sumberdaya ikan supaya tetap produktif dan berkelanjutan. Hal ini disebabkan karena belum adanya sistem informasi yang dapat memprediksi hasil perikanan tangkap berdasarkan pendekatan *data driven* atau data yang dikumpulkan, dianalisis sebagai landasan utama dalam pengambilan keputusan [4].

Prediksi hasil tangkapan ikan merupakan aspek krusial dalam pengelolaan perikanan. Namun, prediksi yang akurat seringkali sulit dilakukan karena banyaknya faktor yang mempengaruhi, seperti iklim, musim, arus laut, dan aktivitas manusia [5] [6]. Meskipun telah digunakan secara luas dengan penggunaan algoritma [7], metode konvensional seperti Regresi Linear memiliki keterbatasan dalam menangani pola sekuensial pada data, yang seringkali ditemui dalam dinamika perubahan hasil tangkapan ikan. Oleh karena itu, pendekatan baru seperti *Recurrent Neural Network* menawarkan potensi untuk meningkatkan akurasi prediksi dengan kemampuannya dalam memahami pola sekuensial pada data deret waktu [8]. Diharapkan bahwa penelitian ini akan memberikan pemahaman yang lebih dalam tentang efektivitas kedua metode dalam memprediksi hasil tangkapan ikan di Laut Sulawesi, sehingga hasilnya dapat menjadi landasan bagi pengembangan kebijakan yang lebih cerdas dalam pengelolaan perikanan di wilayah tersebut, mengarah pada upaya yang lebih berkelanjutan dan berdampak positif bagi masyarakat yang bergantung pada perikanan tangkap.

Permasalahan yang diteliti dalam penelitian ini adalah pertama; prediksi hasil tangkapan ikan yang akurat sulit dilakukan karena banyaknya faktor yang mempengaruhi, seperti iklim, musim, arus laut, dan aktivitas manusia. Serta kedua; penggunaan RNN dalam prediksi hasil tangkapan dalam prediksi hasil tangkapan ikan di wilayah perairan Sulawesi belum banyak dieksplorasi.

Untuk itu tujuan penelitian ini adalah memprediksi produksi hasil tangkapan ikan Cakalang dengan menerapkan metode Regresi Linear dan *Recurrent Neural Network* (RNN), membandingkan dan menganalisis kinerja antara kedua metode tersebut, serta mengidentifikasi faktor-faktor yang berkontribusi pada prediksi hasil tangkapan ikan Cakalang.

II. METODE

A. Waktu dan Tempat Penelitian

Penelitian dilaksanakan selama 9 bulan, dari Maret 2023 sampai dengan Desember 2023, dimulai dengan pengumpulan data primer yakni hasil tangkapan ikan cakalang (*Katsuwonus pelamis*), dilaksanakan pada Balai

Pelabuhan Perikanan Pantai (B-PPP) Tumumpa, Provinsi Sulawesi Utara. Data faktor lingkungan yakni suhu permukaan laut, kadar klorofil, arus laut, dan angin laut diambil dari situs web *Marine Copernicus* dan *Climate Copernicus*. Pengambilan data dilakukan dari bulan Mei 2023 sampai Juli 2023. Pengolahan data dilaksanakan di Laboratorium Rekayasa Perangkat Lunak, Fakultas Teknik, Universitas Sam Ratulangi.

B. Pengambilan Data

Data yang digunakan dalam penelitian ini adalah data hasil tangkapan ikan Cakalang (*Katsuwonus pelamis*) di perairan Laut Sulawesi Utara, dan data faktor lingkungan (suhu permukaan laut, kadar klorofil air laut, arus laut, dan angin laut), seperti terlihat pada Tabel 1.

TABEL 1.
PARAMETER YANG DIANALISIS DALAM PENELITIAN INI.

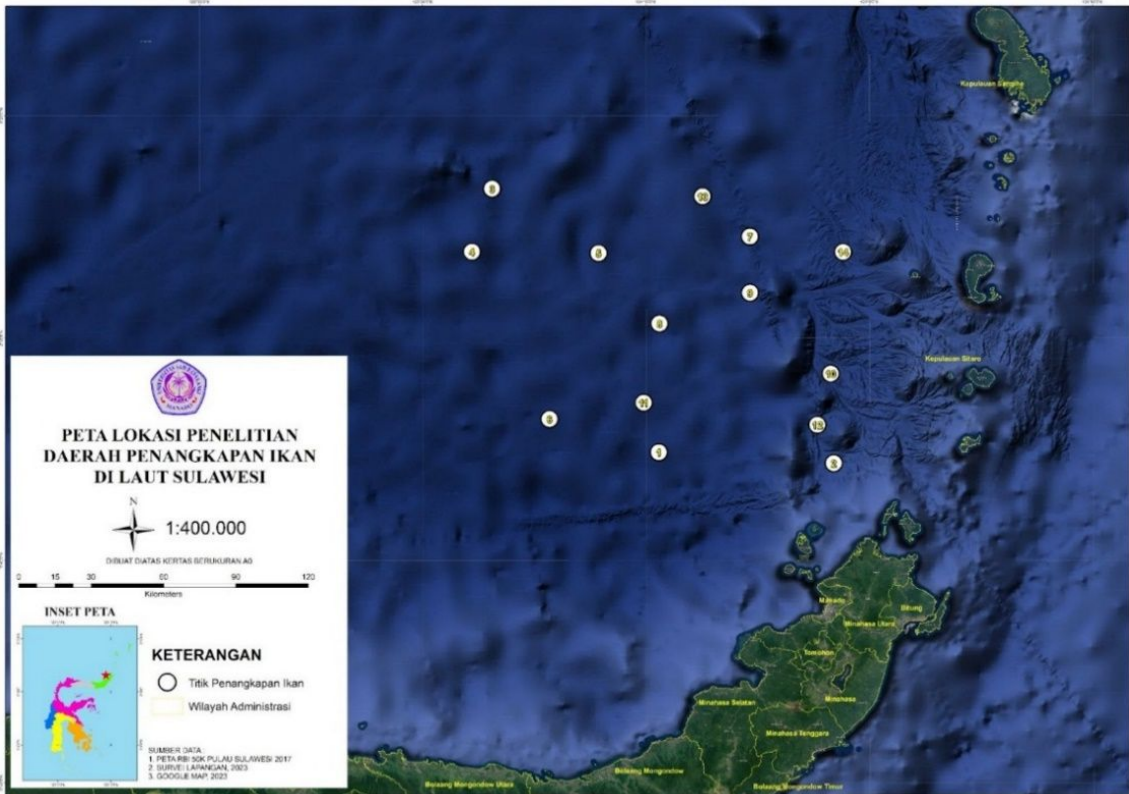
No	Parameter Analisis	Kegunaan
1	Data hasil tangkapan ikan Cakalang (<i>K. pelamis</i>) yang didaratkan di PPP Tumumpa Kota Manado selama 5 tahun (2018 – 2022)	Data yang akan digunakan untuk memprediksi produksi perikanan tangkap di Laut Sulawesi
2	Data titik koordinat daerah penangkapan ikan cakalang (<i>K. pelamis</i>)	Sebagai titik koordinat yang akan dianalisis parameter oseanografi dan hasil tangkapan ikan cakalang (<i>K. pelamis</i>)
3	Data SPL (<i>Global Ocean OSTIA Sea Surface Temperature and Sea Ice Analysis</i>)	Data yang menunjukkan distribusi suhu sebagai variabel bebas dan yang akan dianalisis dengan metode Regresi Linear dan <i>Recurrent Neural Network</i>
4	Data Klorofil (dataset <i>Global Ocean Colour (Copernicus-Glob Colour), Bio-Geo-Chemical</i>)	Data yang menunjukkan distribusi arus sebagai variabel bebas dan yang akan dianalisis dengan metode Regresi Linear dan <i>Recurrent Neural Network</i>
5	Data Arus (<i>Global Ocean Physics Analysis and Forecast</i>)	Data yang menunjukkan distribusi arus sebagai variabel bebas dan yang akan dianalisis dengan metode Regresi Linear dan <i>Recurrent Neural Network</i>
6	Data angin (data <i>ERA5 reanalysis</i>)	Data yang menunjukkan distribusi angin sebagai variabel bebas dan yang akan dianalisis dengan metode Regresi Linear dan <i>Recurrent Neural Network</i>

Data hasil tangkapan ikan Cakalang diambil di kantor Pelabuhan Perikanan Pantai (P3) Tumumpa, Kota Manado, Provinsi Sulawesi Utara. Data yang diambil merupakan data hasil tangkapan yang dilaporkan di Pelabuhan Pantai Tumumpa selama 5 (lima) tahun dari 2018 - 2022.

Data pengukuran faktor lingkungan yang digunakan diunduh dari situs web *Marine Copernicus* dan *Climate Copernicus*. [9], Kedua situs web merupakan bagian dari *Copernicus Programme* yang fokus pada pemantauan dua aspek utama lingkungan, yakni lautan dan iklim. Data yang

diambil dari situs web *Marine Copernicus* adalah suhu permukaan laut (SPL) dan kadar klorofil air laut. Sedangkan dari situs web *Climate Copernicus* diambil data kecepatan arus laut dan kecepatan angin laut.

Data yang diambil merupakan data pengukuran di 14 (empat belas) titik di wilayah Laut Sulawesi (terletak antara lintang 0.8056° hingga 4.325° Utara dan bujur 120° hingga 126° Timur), yang kemudian dihitung nilai rata-rata untuk mendapatkan nilai pengukuran yang mewakili seluruh wilayah Laut Sulawesi (Gambar 1).



Gambar 1. Lokasi pengambilan data variabel lingkungan.

Data SPL yang digunakan adalah *Global Ocean OSTIA Sea Surface Temperature and Sea Ice Analysis*. Data kadar klorofil air laut (mg/m^3). yang digunakan adalah dataset *Global Ocean Colour (Copernicus-GlobColour)*, *Bio-Geo-Chemical*. Data arus laut (m/s). yang digunakan adalah *Global Ocean Physics Analysis and Forecast*. (m/s). Data angin laut yang digunakan adalah data *ERA5 reanalysis*.

C. Pengolahan Data

Data hasil tangkapan ikan Cakalang yang diambil dalam format file excel, sedangkan data pengukuran faktor lingkungan yang diunduh dari situs web *Marine Copernicus* dan *Climate Copernicus* merupakan data dalam format file netCDF. Data-data ini perlu disatukan ke dalam satu format file, *Comma-Separated Values (.csv)*, kemudian disimpan di *Google Drive*

Format netCDF (*Network Common Data Form*) memiliki 2 bagian utama, yakni header yang berisi *metadata* yang menjelaskan isi dari file, seperti nama variabel, dimensi, satuan, dan informasi lainnya yang diperlukan untuk menginterpretasi data (Gambar 2), dan *data array* yang berisi nilai-nilai data sebenarnya yang direpresentasikan dalam format yang terstruktur sesuai dengan dimensi-dimensi yang telah didefinisikan. Untuk mengakses file netCDF, digunakan library *Xarray*.

```
<xarray.Dataset>
Dimensions: (longitude: 30, latitude: 15, time: 1826)
Coordinates:
* longitude (longitude) float32 119.6 119.9 120.1 120.4 ... 126.4 126.6 126.9
* latitude (latitude) float32 4.22 3.97 3.72 3.47 ... 1.47 1.22 0.97 0.72
* time (time) datetime64[ns] 2018-01-01T18:00:00 ... 2022-12-31T18:00:00
Data variables:
  u10 (time, latitude, longitude) float32 ...
  v10 (time, latitude, longitude) float32 ...
Attributes:
Conventions: CF-1.6
history: 2023-07-07 05:09:09 GMT by grib_to_netcdf-2.25.1: /opt/ecmw...
```

Gambar 2. Metadata File NetCDF

Langkah pengolahan data selanjutnya adalah tahapan transformasi data, integrasi data hasil tangkapan dengan data pengukuran faktor lingkungan, dan ekstraksi fitur. Dalam penelitian ini hanya dipilih data waktu penangkapan dan hasil tangkapannya dalam file excel, dengan fungsi *read_excel()* dari library *Pandas*. Setelah data hasil tangkapan ikan dibuka, kolom waktu penangkapan perlu dikonversi tipe datanya menjadi *datetime* dan kolom jumlah hasil tangkapan ikan Cakalang perlu dikonversi tipe datanya menjadi numerik.

Untuk pengukuran faktor lingkungan seperti data arus laut dan angin laut, data yang tersedia merupakan dataset yang terdiri dari dua komponen variabel dataframe, yakni *eastward velocity* (kecepatan ke arah timur) dan *northward velocity* (kecepatan ke arah utara). Kedua komponen dataframe ini perlu digabungkan ke dalam satu dataframe, agar lebih mudah untuk diolah lebih lanjut. Untuk menggabungkan dua data frame, dengan menggunakan fungsi *merge()* dari library *Pandas*. Setelah data netCDF diubah menjadi dataframe yang siap digunakan, kemudian dataframe tersebut diubah dalam format *Comma-Separated Values (.csv)* kemudian disimpan ke *Google Drive*.

Untuk menghitung besaran kecepatan angin dari komponen *eastward & northward velocity*, digunakan teorema pythagoras pada segitiga siku-siku, dengan *northward velocity* sebagai sisi vertikal, dan *eastward velocity* sebagai sisi horizontal (Rumus Pythagoras).

$$V_{total} = \sqrt{(V_{timur})^2 + (V_{utara})^2}$$

1. Exploratory Data Analysis (EDA)

EDA adalah langkah awal yang penting dalam proses analisis data. EDA dapat membantu mendeteksi kesalahan, mengidentifikasi *outlier* dalam kumpulan data, memahami hubungan antar data, menggali faktor-faktor penting,

menemukan pola dalam data, dan memberikan wawasan baru [10]. Dalam tahapan eksplorasi ini diterapkan teknik visualisasi data dan eksplorasi variabel. dengan menggunakan library *Matplotlib*.

Salah satu cara untuk mengetahui distribusi data yang digunakan adalah dengan membuat histogram untuk setiap atribut numerik. dengan menggunakan fungsi *subplots()* dari library *Matplotlib* pada seluruh kumpulan data, dan fungsi ini akan membuat time series plot untuk setiap atribut numerik pada data.

Untuk melihat anomali pada data yang dimiliki, maka dilakukan analisis karakteristik data seperti : mean, median, count, deviasi standar, Q1, Q2, Q3, minimum, dan maximum, dan membandingkannya dengan distribusi data, dengan menggunakan fungsi *describe()*. Untuk analisis hubungan antar variabel digunakan scatter plot untuk hubungan setiap atribut numerik satu sama lain menggunakan fungsi *scatter_matrix()* dari library *Pandas*. *Pandas* memiliki fungsionalitas untuk membuat plot langsung dari data yang ada dalam dataframe atau series, memanfaatkan kemampuan visualisasi yang disediakan oleh *Matplotlib*.

Untuk mengukur hubungan antara variabel hasil tangkapan ikan cakalang dengan variabel lingkungan, digunakan Koefisien Korelasi Pearson (*Pearson's r*), menggunakan fungsi *corr()* dari library *Pandas*. Nilai ini menyatakan bahwa jika koefisien antara variabel mendekati 1, maka terdapat hubungan positif yang kuat antara variabel, dan jika mendekati -1, maka terdapat hubungan negatif yang kuat. Sedangkan jika mendekati 0, maka tidak memiliki hubungan yang signifikan.

D. Pemodelan Data dan Evaluasi Model

Untuk permodelan dan prediksi data, digunakan teknik Regresi Linear dan teknik *Recurrent Neural Network*. Model data dibuat dengan menggunakan library *Scikit-learn*. Khusus untuk *Recurrent Neural Network*, digunakan library *Keras* yang telah disertakan dalam library *Tensorflow*. Keduanya sama-sama digunakan dalam pemodelan Machine Learning, namun *Keras* berfokus pada pemodelan *Neural Network*.

1. Model Regresi Linear

Sebelum data dimodelkan, data dinormalisasikan (rentang nilai di tiap atribut diubah menjadi antara 0 dan 1) dengan *MinMaxScaler* agar nantinya model dapat menggeneralisasi data dengan lebih baik. Ini membantu algoritma Machine Learning untuk tidak terpengaruh oleh perbedaan rentang nilai antar atribut [11]

Untuk memprediksi hasil tangkapan ikan Cakalang dengan teknik Regresi Linear, digunakan semua atribut atau variabel lingkungan sebagai variabel prediktor.

Berdasarkan hasil perhitungan nilai Koefisien Korelasi Pearson, variabel yang memiliki hubungan positif dengan hasil tangkapan ikan Cakalang adalah *sst_dailymean_(C)*, *northward_wind_dailymean_(m/s)*, *eastward_seawater_dailymean_(m/s)*, dan *chlorofil_dailymean_(mg/m³)*.

Setelah variabel prediktor dan target ditentukan, dataset dibagi untuk training dan testing. Proporsi dataset untuk training adalah 75%, dan sisanya untuk testing. Untuk memudahkan pembagian dataset, digunakan fungsi

train_test_split() dari modul *sklearn.model_selection*. Modul ini merupakan bagian dari library *Scikit-learn*. Terlebih dahulu data mengimpor model *LinearRegression* dari modul *sklearn.linear_model*. kemudian, model diinstansiasi menjadi sebuah objek, dan data training dilatihkan ke objek model regresi linear melalui fungsi *fit()*. Setelah model dilatih dengan data training, maka dilakukan prediksi dari data testing dengan menggunakan fungsi *predict()*. Setelah prediksi dilakukan, hasil prediksi dengan data testing digabungkan, dan dilakukan pembalikan normalisasi pada hasil prediksi. Normalisasi yang dilakukan pada data testing sebelum pelatihan model juga dibalikkan.

Evaluasi kinerja model dilakukan dengan menggunakan metrik Koefisien Determinasi (R^2), *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Evaluasi dilakukan dengan menggunakan fungsi metrik yang disediakan oleh modul *sklearn.metrics*.

2. Model Recurrent Neural Network (RNN)

Untuk memprediksi hasil tangkapan ikan Cakalang dengan teknik RNN, langkah pertama adalah dengan menyiapkan dataset, yang meliputi normalisasi, pembagian data untuk training dan testing, dan penentuan variabel input dan output [12].

Yang pertama dilakukan adalah normalisasi dataset (rentang nilai di setiap atribut diubah menjadi antara 0 dan 1) dengan *MinMaxScaler* agar model dapat menggeneralisasi data lebih baik. Ini membantu algoritma RNN untuk tidak terpengaruh oleh perbedaan rentang nilai antar atribut. Setelah itu data dibagi menjadi data training dan testing.

Sama seperti Regresi Linear, penentuan proporsi dataset untuk training adalah 75%, atau 1327 dari total 1770 sampel data, dan sisanya untuk testing. Setelah dibagi menjadi data training dan testing, kedua data ini dibagi lagi menjadi input dan output. Input terdiri dari semua kolom variabel lingkungan, sedangkan output adalah kolom hasil tangkapan Cakalang. Input dibentuk menjadi format 3 dimensi seperti yang bisa diterima oleh format LSTM, yakni [*samples, timesteps, features*].

Untuk arsitektur model RNN, digunakan 1 LSTM dengan 50 neuron sebagai hidden layer dan 1 Dense layer dengan 1 neuron sebagai output layer. Input shape yang diterima model adalah 1 time step dengan 8 atribut factor lingkungan (Gambar 3).

Model dilatih dengan fungsi *fit()*. dengan menggunakan Mean Absolute Error (MAE) sebagai loss function dan Adam optimizer. Model dilatih selama 100 epoch dengan ukuran batch 72. Hal yang harus diatur juga adalah parameter *validation_data* agar dapat mengecek testing loss selama model dilatih.

```
# design network
model = Sequential()
model.add(LSTM(50, input_shape=(train_input.shape[1], train_input.shape[2])))
model.add(Dense(1))
model.compile(loss='mae', optimizer='adam', metrics=["mae"])
```

Gambar 3. Arsitektur Model RNN

Setelah model dilatih, kemudian dilakukan prediksi pada data testing dengan fungsi *predict()*. Sama seperti di metode Regresi Linear, hasil prediksi digabungkan dengan data

testing agar normalisasi yang dilakukan pada kedua data dapat dibalikkan.

Dengan hasil prediksi dan nilai aktual dalam skala aslinya, kinerja model tersebut dapat dievaluasi. Metrik yang digunakan adalah R-Squared Score, Mean Squared Error (MSE), Mean Absolute Error (MAE), dan Root Mean Squared Error (RMSE).

III. HASIL & PEMBAHASAN

A. Exploratory Data Analysis

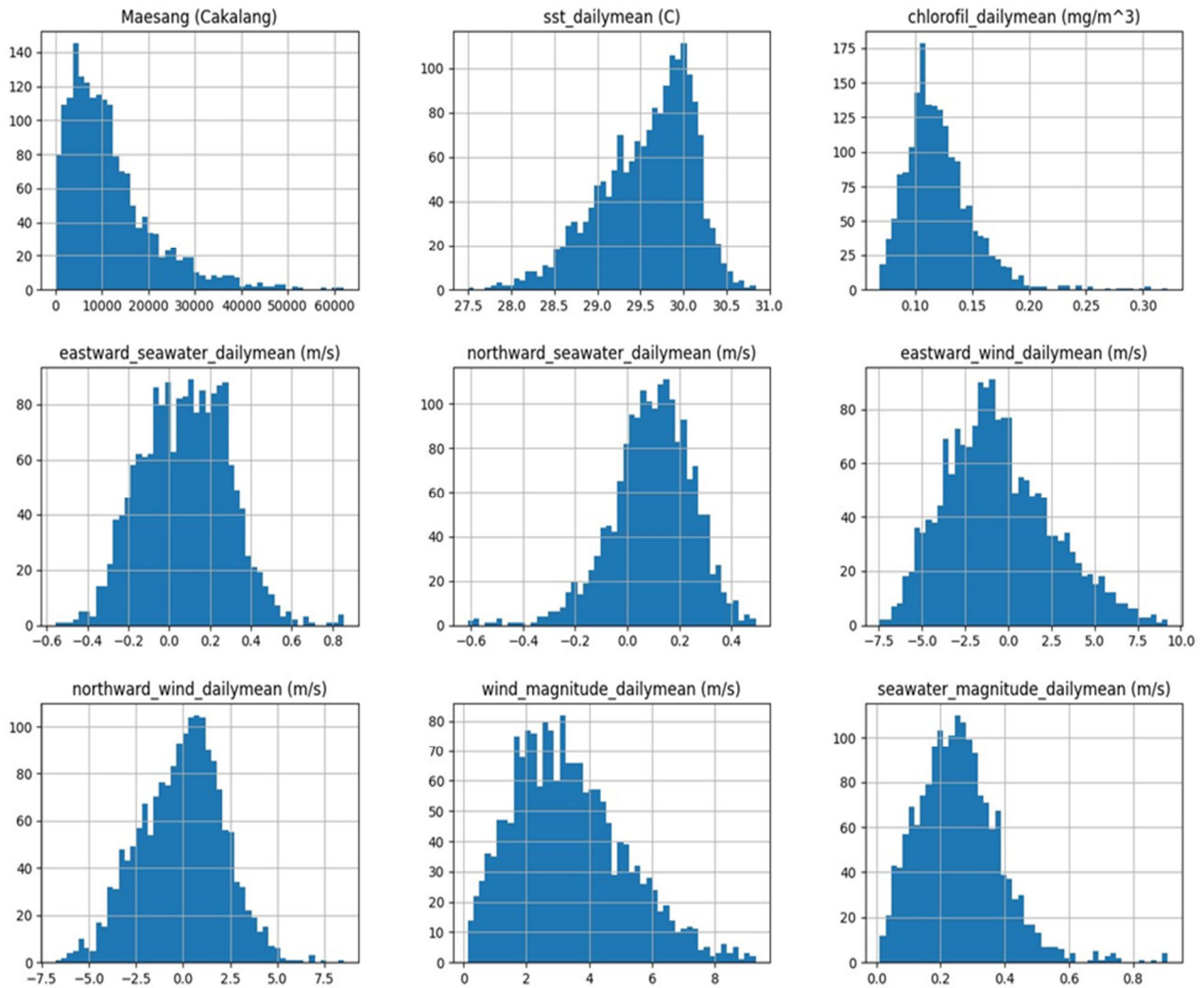
Hasil *Exploratory Data Analysis*, dan hasil analisis dari proses disajikan Gambar 4.

1. Distribusi Data

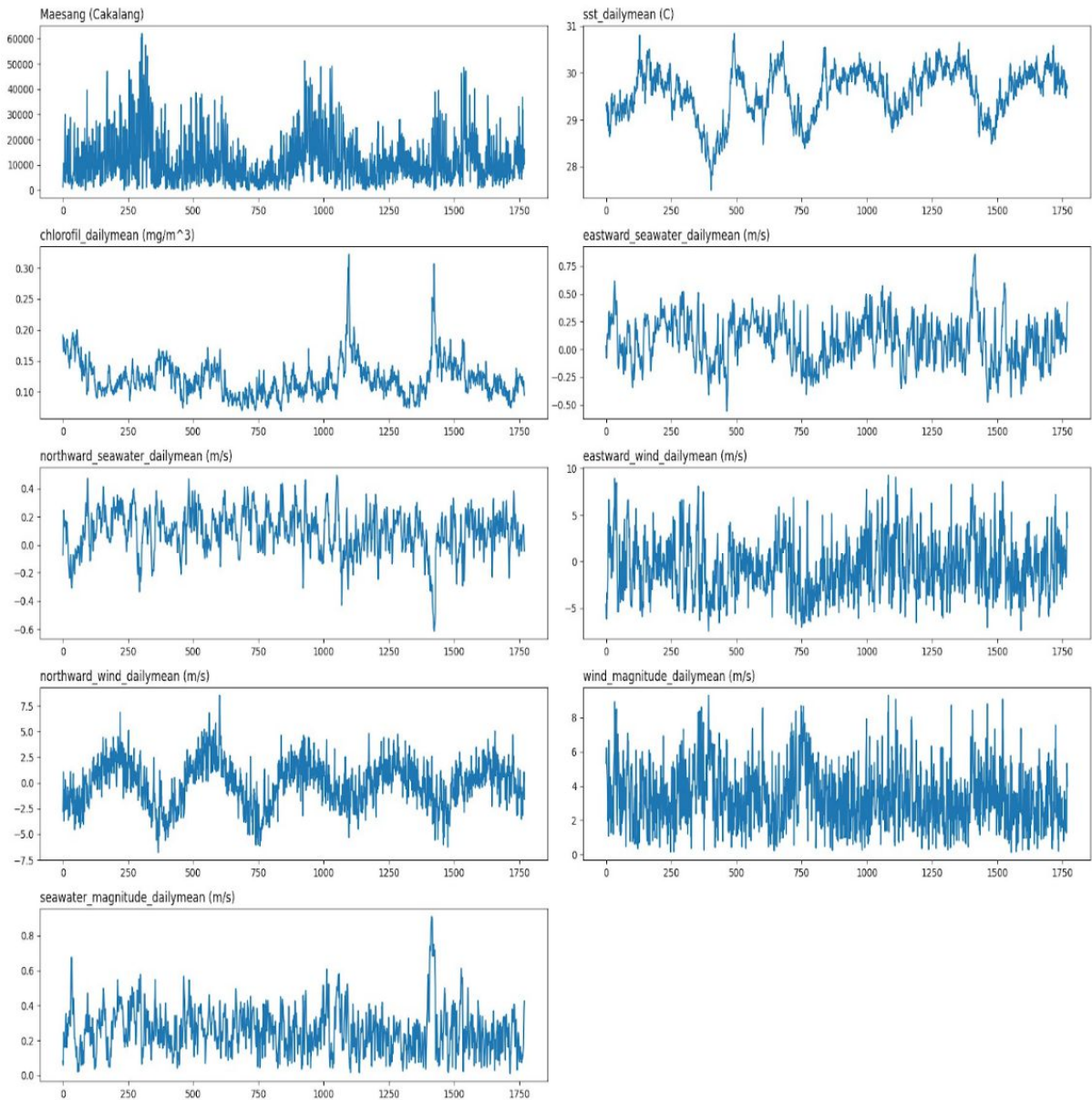
Hasil analisis distribusi data dapat dilihat pada Gambar 4 Berdasarkan kumpulan histogram pada gambar 4, dapat dilihat bahwa ada beberapa atribut yang distribusinya miring dari median masing-masing, baik ke kiri maupun ke kanan, dan yang paling terlihat jelas adalah atribut hasil tangkapan ikan Cakalang. Untuk itu dilakukan transformasi terhadap atribut hasil tangkapan ikan Cakalang agar distribusinya menjadi lebih simetris.

2. Tren Data

Hasil analisis tren data disajikan dalam Gambar 5.



Gambar 4. Visualisasi Distribusi Data



Gambar 5. Visualisasi Tren Data

Berdasarkan *time series* plot pada gambar 5, dapat dilihat bahwa hampir semua variabel memiliki tren naik turun, kecuali variabel SPL yang cenderung di atas, dan variabel kadar klorofil yang cenderung di bawah. Jika kita perhatikan variabel kecepatan angin ke arah utara, terlihat jelas bahwa variabel ini memiliki pola musiman yang paling teratur.

3. Anomali Data

Hasil karakteristik data disajikan dalam Tabel 2. Terlihat bahwa hasil tangkapan rata-rata ikan cakalang adalah 1166.06 kg, nilai SPL berkisar antara 27.49°C, nilai klorofil berkisar antara 0.069 - 0.322, nilai Kecepatan arus rata rata 0.077 m/s arah barat dan 0.557 m/s arah selatan, nilai kecepatan angin rata rata 3.366 m/s arah barat dan 0.26 m/s arah selatan.

TABEL 2
 KARAKTERISTIK DATA

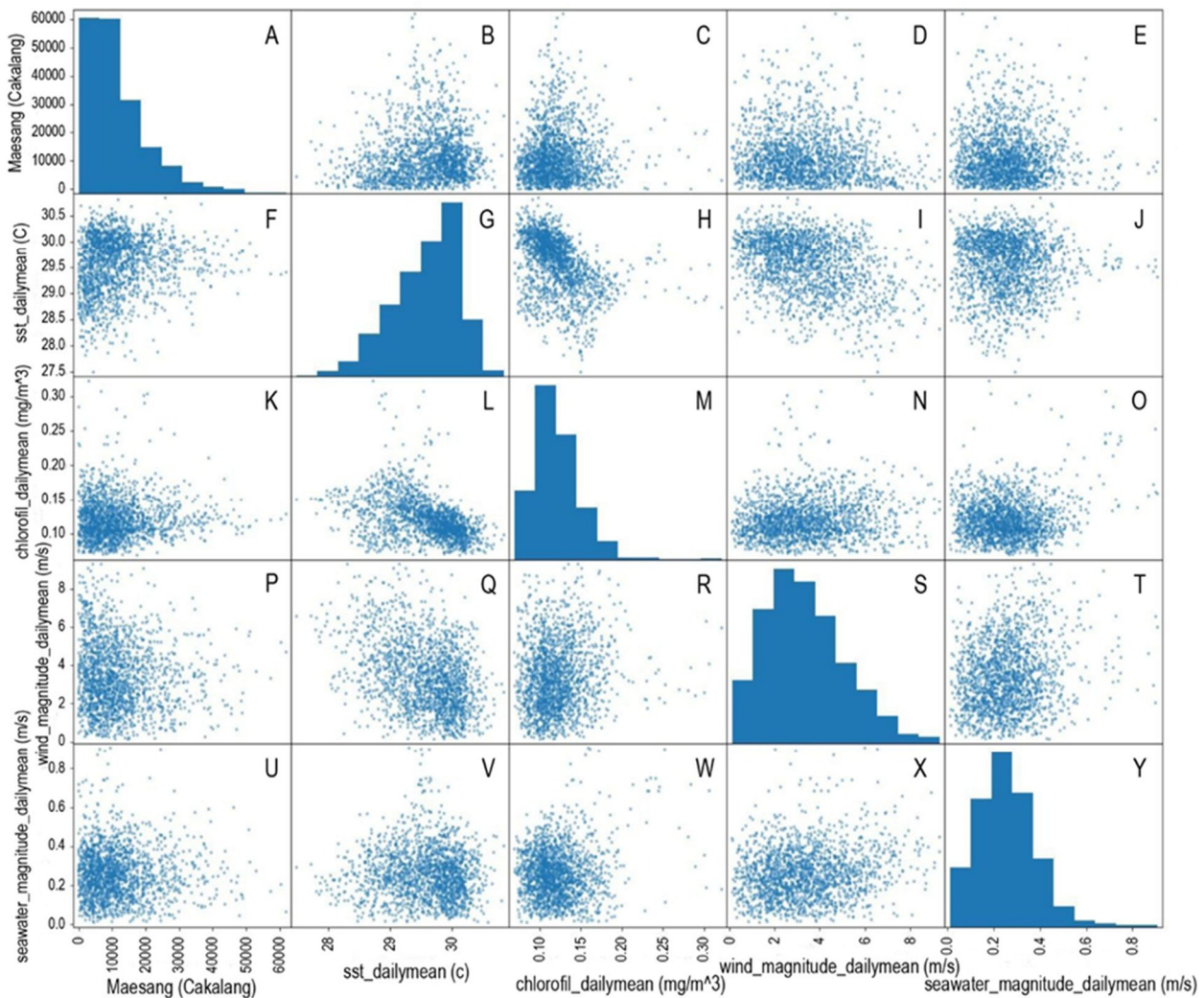
	Ikan Cakalang (kg)	SPL (Co)	Klorofil (mg/m ³)	Kec. Arus Timur	Kec. Arus Utara	Kec angin Timur	Kec angin Utara	Besaran kec. angin Timur	Besaran kec. arus Utara
Jumlah	1770.00	1770.00	1770.00	1770.00	1770.00	1770.00	1770.00	1770.00	1770.00
Rata-rata	1166.063	29.57	0.121	0.077	0.095	-0.557	-0.0659	3.36	0.26
Std	9277.152	0.538	0.029	0.210	0.154	3.048	2.198	1.77	0.13
Min	0.00	27.49	0.069	-0.557	-0.162	-7.426	-6.757	0.13	0.01
25%	4850.00	29.22	0.1018	-0.0747	0.005	-.77	-1.59	2,020	0.167
50%	9490.50	29.66	0.1166	0.0815	0.104	-0.881	0.110	3.156	0.2476
75%	15421.00	29.98	0.1353	0.233	0.200	1.367	1.444	44.454	0.331
Max	61992.00	30.84	0.3218	0.855	0.492	9.251	8.513	9.30	0.91

Jika dilihat dari karakteristik data pada Tabel 2, terdapat anomali pada atribut hasil tangkapan ikan Cakalang. Pada kuartil pertama (25%), kedua (50%), dan ketiga (75%), selisih antar kuartil tidak terlalu jauh, yakni kurang lebih 5000 kg. Namun selisih antara kuartil ketiga dengan nilai maksimum terpaut sangat jauh, yakni kurang lebih 46000 kg. Tidak mengherankan jika dibandingkan dengan histogram atribut hasil tangkapan ikan Cakalang, dimana terlihat bahwa distribusi data tersebut condong ke kiri (Gambar 6). Untuk itu seperti yang dijelaskan dibagian distribusi data, jika diperlukan, dapat dilakukan transformasi terhadap atribut ini untuk menormalkan distribusi data.

4. Hubungan Antar Variabel

Hasil analisis hubungan antar variabel yang dianalisis dapat dilihat pada Gambar 6.

Berdasarkan gambar 6, dapat dilihat bahwa ada hubungan signifikan diantara beberapa variabel lingkungan yakni antara variabel SPL dan variabel klorofil yang bersifat negatif (gambar 6 H dan L), artinya jika nilai suhu permukaan naik maka nilai klorofil akan turun. dan hubungan negatif juga terlihat antara variabel SPL dan Kecepatan angin (gambar 6 Q). Hal yang sama juga ditemukan dalam studi tentang hubungan antara variabel SPL dan klorofil dari data citra Sentinel-3 OLCI dan SLST, dimana kedua variabel tersebut menunjukkan nilai korelasi negatif [13]. Sedangkan untuk korelasi antara variabel yang lain kurang jelas terlihat.



Gambar 6. Visualisasi Hubungan Antar Variabel

Namun hubungan antara variabel lingkungan yang dianalisis dan hasil tangkapan ikan cakalang, terlihat relatif rendah, yakni hanya dapat dijelaskan kurang dari 25% selebihnya karena pengaruh faktor lain. Nilai hubungan linier antara variabel hasil tangkapan ikan cakalang dengan variabel lingkungan dianalisis dari nilai koefisien Korelasi Pearson (*Pearson's r*), menggunakan fungsi *corr()* dari library *Pandas*. Nilai ini menyatakan bahwa jika koefisien antara variabel mendekati 1, maka terdapat hubungan positif yang kuat antara variabel, dan jika mendekati -1, maka terdapat hubungan negatif yang kuat. Sedangkan jika

mendekati 0, maka tidak memiliki hubungan yang signifikan.

Hubungan antara parameter hasil tangkapan cakalang dan variabel lingkungan yang dinyatakan dalam koefisien korelasi Pearson, dapat dilihat pada Tabel 3.

TABEL 3
NILAI KORELASI ANTAR VARIABEL YANG DIANALISIS

Variabel	Nilai Korelasi
Ikan Cakalang (kg)	1.00
SPL (°C)	0.117
Kec. angin ke utara (m/s)	0.115
Kec. arus ke timur (m/s)	0.07
Klorofil (mg/m ³)	0.04
Kec. arus ke utara (m/s)	-0.006
Kec. angin ke timur (m/s)	-0.011
Besaran kec. arus	-0.017
Besaran kec. angin	-0.087

Dari Tabel 3, dapat dilihat bahwa hubungan antara hasil tangkapan ikan cakalang dan variabel oseanografi seperti suhu, kecepatan angin ke utara, kecepatan arus ke timur, dan klorofil adalah positif, sedangkan hubungan dengan variabel kecepatan arus ke utara, kecepatan angin ke timur, besaran kecepatan angin dan arus adalah negatif.

Hasil analisis menunjukkan bahwa hubungan antara jumlah hasil tangkapan ikan cakalang yang didaratkan di PPP Tumumpa dengan faktor lingkungan tergolong rendah yakni dengan faktor SPL dan kecepatan angin ke utara sebesar 11%, besaran kecepatan angin 8%, kecepatan arus ke timur 7%, klorofil 4%, besaran kecepatan arus dan kecepatan angin ke timur 1%, dan kecepatan arus ke utara hanya 0.6%. Hal ini menjelaskan ada pengaruh faktor lain yang belum diteliti dalam penelitian ini.

Hasil penelitian dengan menggunakan *Backpropagation Neural Network*, menemukan bahwa suhu udara dinilai kurang mempengaruhi terhadap hasil akhir peramalan hasil tangkapan ikan [14].

B. Analisis Hasil dan Evaluasi Model

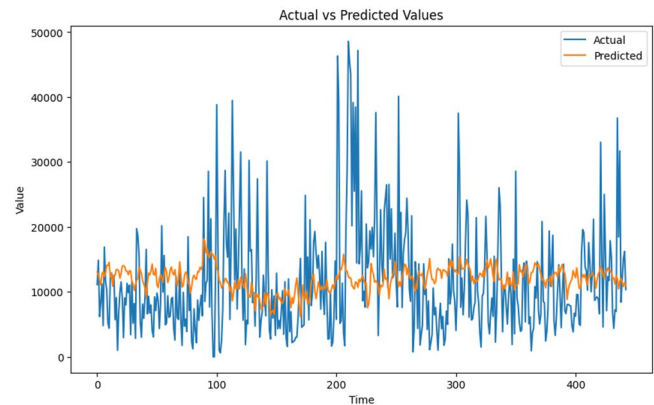
Berikut adalah hasil pemodelan, dan analisis tentang hasil dari pemodelan ini.

1. Regresi Linear

Setelah melakukan prediksi pada data tes, maka perbandingan hasil prediksi data tes dengan data tes asli seperti terlihat pada Tabel 4. dan Gambar 7.

TABEL 4
HASIL PREDIKSI NILAI AKTUAL DAN PREDIKSI UNTUK MODEL REGRESI LINEAR

No	Nilai Aktual	Nilai Prediksi
0	11120.0	13057.48
1	14870.0	12676.20
2	6199.0	11757.47
3	8410.0	11057.51
4	11240.0	12808.64
5	4800.0	13039.57
6	16902.0	12346.56
7	11647.0	12930.72
8	10250.0	14094.39
9	5150.0	14078.88
10	4414.0	14551.91
11	12486.0	12290.52
12	12906.0	12825.45
13	11825.0	10868.27
14	12479.0	12614.33



Gambar 7. Plot Hasil Prediksi Model Regresi Linear

Berikut merupakan kinerja model Regresi Linear yang diukur dengan berbagai metrik, yakni Koefisien Determinasi (R^2), *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE) (Tabel 5).

TABEL 5
HASIL EVALUASI KINERJA MODEL REGRESI LINEAR

Koefisien Determinasi (R^2)	-0.04908941733804384
Mean Absolute Error (MAE)	6527.0086
Mean Square Error (MSE)	75986790.8
Root Mean Square error (RMSE)	8717.0402

Dapat dilihat bahwa nilai Koefisien Determinasi berada di angka minus 4 persen. Dan nilai MAE berada disekitar 6500, yang berarti rata-rata selisih antara nilai prediksi dengan nilai sebenarnya ada di sekitar 6500 kilogram. Nilai ini masih terlalu besar, dan bisa berarti:

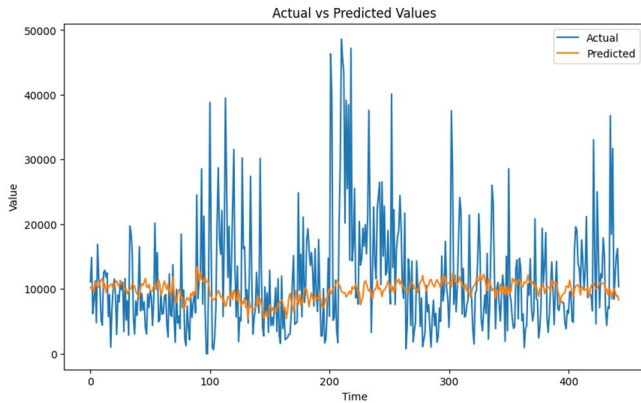
- 1) Variabel prediktor yang digunakan (SPL, Klorofil, Angin Utara, dan Arus Timur) memiliki pengaruh yang kurang signifikan terhadap hasil tangkapan ikan Cakalang.
- 2) Model belum bisa menggeneralisasi data dengan baik. Hal ini mungkin disebabkan oleh adanya pengaruh faktor lain yang belum dimasukkan sebagai variabel dalam penelitian.

2. Recurrent Neural Network (RNN)

Setelah melakukan prediksi pada data tes, maka dapat dilihat perbandingan hasil prediksi data tes dengan data tes asli pada Tabel 6 dan Gambar 8.

TABEL 6
HASIL PREDIKSI NILAI AKTUAL DAN PREDIKSI UNTUK MODEL RNN

No	Nilai Aktual	Nilai Prediksi
0	11120.0	10256.20
1	14870.0	9994.53
2	6199.0	10296.09
3	8410.0	9536.72
4	11240.0	11016.96
5	4800.0	10984.82
6	16902.0	10206.63
7	11647.0	10346.43
8	10250.0	11109.96
9	5150.0	11345.72
10	4414.0	10108.00
11	12486.0	11669.08
12	12906.0	10842.07
13	11825.0	9472.15
14	12479.0	10509.39



Gambar 8. Plot Hasil Prediksi Model RNN

Hasil kinerja model RNN yang diukur dengan metrik yang sama digunakan di model Regresi Linear, yakni Koefisien Determinasi (R^2), *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE) dapat dilihat pada Tabel 7.

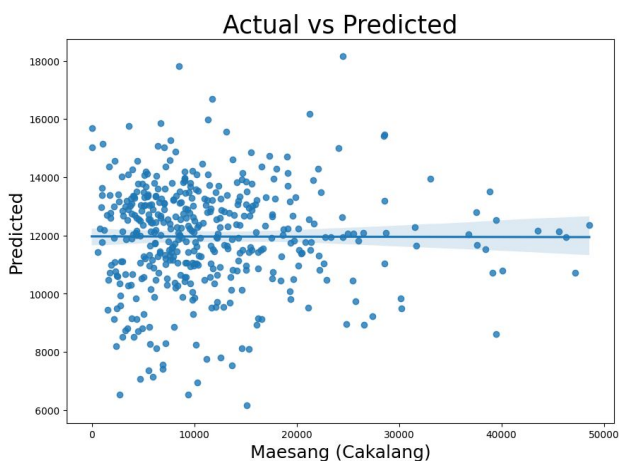
TABEL 7
 HASIL EVALUASI KINERJA MODEL RNN

Koefisien Determinasi (R^2)	-0.06476842919493353
Mean Absolute Error (MAE)	6054.228
Mean Square Error (MSE)	77122450.0
Root Mean Square Error (RMSE)	8781.938

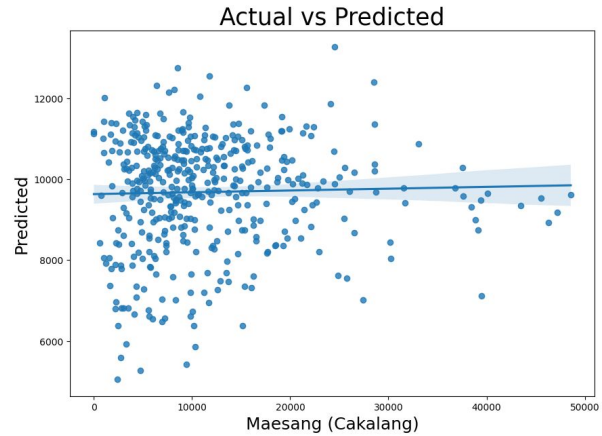
Dapat dilihat bahwa nilai Koefisien Determinasi berada di angka minus 6 persen. Dan nilai MAE berada disekitar 6000, yang berarti rata-rata selisih antara nilai prediksi dengan nilai sebenarnya ada di sekitar 6000 kilogram. Nilai ini masih cukup besar.

Jika dibandingkan hasil evaluasi kedua model, kedua model memiliki nilai Koefisien Determinasi yang buruk, yakni minus 4 persen, dengan model RNN yang memiliki nilai minus 6 persen. Namun nilai MAE model RNN masih lebih baik, yakni 6000 kilogram dibandingkan dengan model Linear yang memiliki nilai MAE 6500 kilogram. Namun hasil evaluasi kedua model diatas menunjukkan bahwa kedua model memiliki kinerja yang belum memuaskan.

Hal ini dapat dilihat juga pada plot sebar nilai aktual vs prediksi (Gambar 9 dan 10) terlihat untuk menggambarkan seberapa baik kinerja model. Untuk model ideal, distribusi titik-titiknya harus lebih dekat ke garis diagonal.



Gambar 9. Distribusi nilai aktual vs prediksi untuk model regresi linear



Gambar 10. Distribusi nilai aktual vs prediksi untuk model RNN

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

1. Variabel prediktor yang digunakan (SPL, Klorofil, Angin, dan Arus) memiliki pengaruh yang kurang signifikan (kurang dari 50%) terhadap hasil tangkapan ikan Cakalang.
2. Performa kedua model, baik Linear maupun RNN masih belum memuaskan dalam memprediksi hasil tangkapan ikan Cakalang. Hal ini dibuktikan dengan hasil evaluasi kedua model yang memiliki nilai Koefisien Determinasi yang buruk, dan nilai MAE yang besar.

B. Saran

1. Perlu mengumpulkan data yang lebih banyak, dan semua data yang digunakan, baik hasil tangkapan maupun faktor lingkungan sebaiknya berasal dari sumber data yang sama, agar data yang digunakan selaras dan tidak menyulitkan proses pra pemrosesan dan transformasi data.
2. Perlu dilakukan eksplorasi data atau penelitian lanjutan untuk menemukan variabel lain yang memiliki pengaruh yang lebih signifikan terhadap hasil tangkapan ikan Cakalang.
3. Masih perlu dilakukan pra pemrosesan dan transformasi pada data yang digunakan, agar memudahkan model dalam menggeneralisasi data.
4. Perlu melakukan *Hyperparameter Tuning* dan optimasi untuk meningkatkan performa model.

V. KUTIPAN

- [1] Simbolon, D. 2019. Daerah Penangkapan ikan. IPB Press. 246 hal
- [2] https://statistik.kkp.go.id/home.php?m=prod_ikan_laut_kab (2019).Statistik KKP, Produksi Perikanan Tangkap Laut
- [3] Dahuri, R., 2007. Membenahi sistem manajemen perikanan tangkap. Majalah Samudra edisi 5-. V. 37-39.
- [4] Pratama G.B., T.W. Nurani, Dan R. I. Wahyu, 2020. Rancang Bangun Sistem Basis Data Untuk Menunjang Pengelolaan Perikanan Tuna Cakalang Dan Tongkol Di Perairan Natuna. Akuatika Indonesia, Vol 5. N0.2. 77-85
- [5] Mann, K. H. dan J. R. N. Lazier. 2006. *Dynamic of Marine Ecosystem, Biological- Physical Interaction in the Ocean, Third Edition*. Canada : Blackwell Publishing
- [6] Laevastu, T dan M. L. Hayes. 1981. *Fisheries Oceanography and Ecology*. Fishing News Books Ltd. England. 199 p
- [7] Campbell J, Antonie D, Armstrong R, Arrigo K, Balch W, Barber R, Behrenfeld M, Bidigare R, Bishop J, Carr ME, Esaias W, Falkowski P, Hoepffner N, Iverson R, Kiefer D, Lohrenz S, Marra J, Morel A, Ryan J, Vedernikov V, Waters K, Yentsch C, Yoder J. 2002. Comparison of algorithm for estimating ocean primary production from

- surface chlorophyll, temperature and irradiance. Global biogeochemical cycles. Vol 16. No 3,1035.
- [8] Nanang Hermawan, “Aplikasi Model *Recurrent NeuralNetwork*,” 2014.
- [9] <https://climate.copernicus.eu/climate-reanalysis> untuk daerah perairan.
- [10] Makhiya. S.K dan U. Ahmed, 2020.Hands-On Exploratory Data Analysis with Python_Perform EDA technique,.
- [11] Asmoro, T., Musyafa, A., & Fahrudin, A. (2018). Predicting Fishery Production in Indonesia using Linear Regression Model. *Aquaculture, Aquarium, Conservation & Legislation*, 11(1), 197-204.
- [12] Alifah, N., Permasari, A. E., & Rustiawan, W. (2020). Predicting Fisheries Production using Recurrent Neural Network in Indonesia. *IOP Conference Series: Materials Science and Engineering*, 725(1), 012007.
- [13] Devita Ningrum, M. Zainuri dan R. Widiaratih 2022. Variabilitas Bulanan Klorofil-A Dan Suhu Permukaan Laut pada Perairan Teluk Rembang Dengan Menggunakan Citra Sentinel-3. *Indonesian Journal of Oceanography (IJOCE)* [May] [2022] Vol 04 No 02 : 88 – 96.
- [14] Razak, M.A. &E. Riksakomara. 2017. Peramalan jumlah produksi ikan dengan menggunakan backpropagation neural network. *Jurnal Teknik ITS*, 6 (1), 142-148.



Theofilio Kamang Tarumingkeng lahir di Manado, 5 Oktober 2001, anak dari Adrie Tarumingkeng (ayah) dan Wilhelmina Patty (Ibu). Penulis berdomisili di Kelurahan Malalayang Satu, Kecamatan Malalayang, Kota Manado. Perjalanan pendidikan penulis dimulai dari SD Kristen Eben Haezar 02 Manado dan lulus pada tahun 2013. Kemudian dilanjutkan di SMP Kristen Eben Haezar 2 Manado dan lulus pada tahun 2016. Pendidikan Menengah Atas ditempuh di SMA Kristen Eben Haezar Manado dan Lulus pada tahun 2019. Kemudian penulis melanjutkan pendidikan S1 di salah satu perguruan tinggi di Sulawesi Utara yaitu Universitas Sam Ratulangi dengan mengambil Program Studi Informatika, Jurusan Elektro, Fakultas Teknik.