



KLASIFIKASI *WHALE SIGHTING ZONES* UNTUK PAUS BUNGKUK BERDASARKAN FAKTOR LINGKUNGAN BERBASIS ALGORITMA *TREE*

FAHMI NOOR FIQRI



**PROGRAM STUDI MAGISTER ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Klasifikasi *Whale Sighting Zones* untuk Paus Bungkuk berdasarkan Faktor Lingkungan berbasis Algoritma *Tree*” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2025

Fahmi Noor Fiqri
G6501231008



RINGKASAN

FAHMI NOOR FIQRI. Klasifikasi *Whale Sighting Zones* untuk Paus Bungkuk berdasarkan Faktor Lingkungan berbasis Algoritma *Tree*. Dibimbing oleh KARLISA PRIANDANA dan MEDRIA KUSUMA DEWI HARDHIENATA.

Paus memiliki peranan ekologi yang vital dan memiliki nilai ekonomi dan sosial yang tinggi. Sebagai hewan mamalia terbesar di dunia, paus berkontribusi sebagai agen yang mendorong produksi oksigen di atmosfer, mengurangi karbon dioksida di udara melalui sekuestrasi karbon, dan membantu fertilisasi ekosistem kelautan. Secara ekonomis, pada tahun 2008, sektor ekowisata pengamatan paus memiliki potensi nilai ekonomi tahunan sebesar €148 juta.

Karena vitalnya peranan paus dalam menjaga ekologi dan kesejahteraan ekonomi masyarakat pesisir, paus merupakan salah satu hewan yang dilindungi dan konservasi paus merupakan salah satu bagian dari pembangunan berkelanjutan (SDG) nomor 14 *Life Below Water*. Beberapa usaha konservasi paus yang telah dilakukan berkaitan dengan pendeteksian dan pelacakan paus agar dapat terhindar dari aktivitas manusia. Telah dikembangkan berbagai model *machine learning* untuk pendeteksian paus mulai dari analisis sinyal audio dan radar, foto jarak dekat melalui *smartphone*, UAV/*drone*, dan helikopter, hingga pemanfaatan citra satelit (*remote sensing*) untuk mendeteksi pergerakan paus.

Pemanfaatan data lingkungan untuk memodelkan zona kemunculan paus dapat memberikan pemahaman baru mengenai dinamika lingkungan dan habitat paus sekaligus menjadi salah satu metode untuk memetakan lokasi dengan potensi kemunculan paus. Informasi tersebut kemudian dapat digunakan untuk membuat kebijakan berkaitan dengan konservasi maupun eksploitasi pada sektor ekowisata. Pada penelitian ini digunakan data kemunculan paus yang mencakup wilayah perairan di laut Hindia. Wilayah penelitian kemudian dibagi menjadi dua wilayah perairan, yaitu wilayah Afrika dan Australia. Selain itu, penelitian ini menggunakan data lingkungan statis yaitu data kedalaman laut atau batimetri dan data lingkungan dinamis yaitu sifat geofisika seperti temperatur dan tekanan air laut serta sifat biogeokimia seperti konsentrasi klorofil-a dan mineral terlarut lainnya. Total banyaknya variabel yang digunakan untuk pemodelan adalah 25 variabel.

Zona kemunculan paus pada penelitian ini didefinisikan sebagai kisi dengan panjang sisi $0,3^{\circ}$ yang tersebar di permukaan laut di wilayah Afrika dan Australia. Untuk setiap zona yang dibuat, dihitung nilai rata-rata untuk setiap variabel sehingga data citra dapat dikonversi menjadi tabel menggunakan metode *zonal statistics*. Setelah dataset dibuat, data dibagi menjadi data latih dan uji dengan proporsi 80% dan 20%. Pembangunan model prediksi zona kemunculan paus dilakukan dengan membandingkan empat algoritma *machine learning* berbasis *tree*, yaitu *decision tree*, *random forest*, *CatBoost*, dan *XGBoost*. Evaluasi model dilakukan menggunakan metode *k-fold cross-validation* dengan nilai $k = 10$ dan ukuran kinerja model yang digunakan adalah *Matthew correlation coefficient* (MCC), akurasi, *precision*, *recall*, dan *F1 score*. Selain itu, dilakukan tiga eksperimen yaitu pemodelan dengan seluruh data latih, pemodelan dengan data latih yang sudah diseimbangkan distribusi kelasnya, dan pembuatan model dengan seleksi variabel untuk membuat model dengan input variabel yang efektif.

Eksperimen pertama dan kedua menunjukkan bahwa model *XGBoost* merupakan model yang memiliki konsistensi kinerja terbaik dengan skor MCC tertinggi sebesar 0,99450 setelah dievaluasi menggunakan data uji. Eksperimen ketiga berhasil mengidentifikasi bahwa model *XGBoost* dapat menghasilkan kinerja yang baik dengan sedikit penurunan skor MCC menjadi 0,99410 dengan menggunakan lima variabel dengan nilai *gain importance* tertinggi. Melalui analisis kontribusi variabel menggunakan metode *Shapley Additive Explanations (SHAP)*, berhasil diidentifikasi peringkat lima faktor lingkungan yang memiliki kontribusi signifikan terhadap kemunculan paus yaitu tekanan air pada dasar laut, jarak zona ke tepi pantai, kemiringan relief pada dasar laut, fosfat terlarut, dan ketinggian permukaan laut.

Kata kunci: klasifikasi, faktor lingkungan, paus bungkuk, *remote sensing*, kemunculan paus



SUMMARY

FAHMI NOOR FIQRI. Humpback Whale Sighting Zones Classification based on Environmental Factors using Tree-based Algorithms. Supervised by KARLISA PRIANDANA and MEDRIA KUSUMA DEWI HARDHIENATA.

Whales play a vital ecological role and hold significant economic and social value. As the largest mammals on Earth, whales contribute to atmospheric oxygen production, reduce atmospheric carbon dioxide through carbon sequestration, and support the fertilization of marine ecosystems. Economically, in 2008, the whale-watching ecotourism sector was estimated to have an annual economic value of €148 million.

Given their essential role in maintaining ecological balance and supporting the livelihoods of coastal communities, whales are protected species, and whale conservation is an integral part of Sustainable Development Goal (SDG) 14: Life Below Water. Various conservation efforts have focused on detecting and tracking whales to minimize harmful human interactions. Numerous machine learning models have been developed for whale detection, ranging from audio and radar signal analysis, close-range imagery via smartphones, UAVs/drones, and helicopters, to satellite-based remote sensing for tracking whale movement.

Leveraging environmental data to model whale sighting zones offers valuable insights into habitat dynamics and can serve as a method for identifying potential whale occurrence locations. Such information can inform policy-making in conservation and ecotourism development. This study utilizes whale sighting data covering the Indian Ocean region, which is divided into two research areas: African waters and Australian waters. The study incorporates both static environmental data—specifically bathymetry—and dynamic environmental data, including geophysical properties (e.g., sea temperature and pressure) and biogeochemical properties (e.g., chlorophyll-a concentration and other dissolved minerals). A total of 25 variables are used in the modeling process.

In this study, a whale sighting zone is defined as a grid cell with a side length of 0.3° , distributed across the ocean surface in the African and Australian regions. For each zone, the mean value of every environmental variable is computed, allowing image data to be converted into tabular form using the zonal statistics method. After constructing the dataset, the data are split into training and testing sets in an 80:20 ratio. The predictive modeling of whale sighting zones is conducted by comparing four tree-based machine learning algorithms: Decision Tree, Random Forest, CatBoost, and XGBoost. Model evaluation is performed using 10-fold cross-validation, with performance metrics including the Matthews Correlation Coefficient (MCC), accuracy, precision, recall, and F1-score. Three experiments are conducted: (1) modeling with the full training set, (2) modeling with class-balanced training data, and (3) feature selection to create a more efficient model using fewer input variables.

The first and second experiments show that XGBoost consistently delivers the best performance, achieving the highest MCC score of 0.99450 on the test data. The third experiment demonstrates that the XGBoost model can still perform well—with only a slight reduction in MCC to 0.99410—when using only the five most important variables, as determined by gain importance. Through variable

contribution analysis using the SHapley Additive Explanations (SHAP) method, the five most influential environmental factors contributing to whale sightings are identified as seafloor water pressure, distance to the coastline, seafloor slope, dissolved phosphate concentration, and sea surface height.

Keywords: classification, environmental factors; humpback whale, remote sensing, whale sightings

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2025
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.



KLASIFIKASI *WHALE SIGHTING ZONES* UNTUK PAUS BUNGKUK BERDASARKAN FAKTOR LINGKUNGAN BERBASIS ALGORITMA *TREE*

FAHMI NOOR FIQRI

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister pada
Program Studi Ilmu Komputer

**PROGRAM STUDI MAGISTER ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Judul Tesis : Klasifikasi *Whale Sighting Zones* untuk Paus Bungkuk
berdasarkan Faktor Lingkungan berbasis Algoritma *Tree*
Nama : Fahmi Noor Fiqri
NIM : G6501231008

Disetujui oleh

Pembimbing 1:

Dr. Karlisa Priandana, S.T., M.Eng.
NIP 19851121 201212 2 002



Pembimbing 2:

Dr. Medria Kusuma Dewi Hardhienata, S.Komp.
NIP 19860822 202012 2 001



Diketahui oleh

Ketua Program Studi:

Irman Hermadi, S.Kom, M.S., Ph.D.
NIP 19750311 200604 1 009



Dekan Sekolah Sains Data, Matematika, dan Informatika

Prof. Dr. Ir. Agus Buono, M.Si., M.Kom.
NIP 1966070 2199302 1 001



Tanggal Ujian:
28 Juli 2025

Tanggal Lulus:

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan ini adalah bagian dari geoinformatika dan konservasi biodiversitas maritim dengan judul “Klasifikasi *Whale Sighting Zones* untuk Paus Bungkuk berdasarkan Faktor Lingkungan berbasis Algoritma *Free*”.

Terima kasih penulis ucapkan kepada:

Ayah, ibu, dan adik-adikku yang selalu memberikan doa dan dukungan.

Dr. Karlisa Priandana, S.T., M.Eng. selaku dosen pembimbing utama dan Dr. Medria Kusuma Dewi Hardhienata, S.Komp. selaku dosen pembimbing kedua yang telah memberikan bimbingan, arahan, dan saran dalam proses penelitian tesis.

5. Irman Hermadi, S.Kom, M.S., Ph.D. selaku Ketua Program Studi S-2 Ilmu Komputer.

4. Seluruh dosen S-2 Ilmu Komputer atas semua ilmu, pengetahuan, dan pengalaman selama proses perkuliahan hingga penelitian tesis ini selesai.

5. Teman-teman seperjuangan S-2 Ilmu Komputer Kelas Karyawan angkatan 2023: Rihadatul, Ujang, Yan, Atik, Rosyid, Yudhya, Iyan, Fandy, Badrotul, Vicky, Teguh, Ibrahim, Reslan, Tresna, Tisna, Alfian, Naufal, Zam-Zam dan Nurafifah.

6. Seluruh pihak yang tidak dapat disebutkan satu per satu yang telah memberikan kerja sama dan dukungan sehingga tesis ini dapat diselesaikan.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Agustus 2025

Fahmi Noor Fiqri

DAFTAR ISI

DAFTAR TABEL	xii
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
1.4 Manfaat	3
II TINJAUAN PUSTAKA	4
2.1 Paus Bungkok	4
2.2 <i>Decision Tree</i>	4
2.3 <i>Random Forest</i>	5
2.4 <i>Gradient Boosting</i>	7
2.5 <i>XGBoost</i>	8
2.6 <i>CatBoost</i>	9
2.7 <i>Shapley Additive Explanations (SHAP)</i>	11
III METODE	12
3.1 Area Studi dan Sumber Data	12
3.2 Peralatan Penelitian	13
3.3 Tahapan Penelitian	14
IV HASIL DAN PEMBAHASAN	20
4.1 Zona Kemunculan Paus	20
4.2 Evaluasi Kinerja Model	20
4.3 Analisis Kontribusi Variabel	23
V SIMPULAN DAN SARAN	26
5.1 Simpulan	26
5.2 Saran	26
DAFTAR PUSTAKA	27
LAMPIRAN	34
RIWAYAT HIDUP	38



DAFTAR TABEL

1	Data yang digunakan pada penelitian	12
2	Keterangan variabel dataset titik kemunculan paus	13
3	Keterangan variabel dataset <i>Global Ocean Physics</i>	13
4	Keterangan variabel dataset <i>Global Ocean Biogeochemistry</i>	13
5	Keterangan variabel dataset yang diturunkan dari analisis medan	15
6	Komponen pada <i>confusion matrix</i>	19
7	Karakteristik zona kemunculan paus	20
8	Kinerja model klasifikasi pada eksperimen pertama	21
9	Kinerja model klasifikasi pada eksperimen kedua	21
10	Tabulasi silang banyaknya kemunculan paus per benua	22
11	Uji multikolinearitas pada lima variabel terpilih	23

DAFTAR GAMBAR

1	Visualisasi Metode <i>Decision Tree</i>	5
2	Visualisasi Metode <i>Random Forest</i> .	6
3	Visualisasi Metode <i>Bagging</i> dan <i>Boosting</i> .	7
4	Visualisasi <i>tree</i> biasa dan <i>tree</i> simetris.	10
5	Tahapan penelitian	14
6	Tahapan penelitian.	14
7	Tren musiman kemunculan paus pada tahun 2021 s.d. 2023	15
8	Visualisasi proses delineasi zona kemunculan paus	16
9	Wilayah <i>pseudo-absence</i> pengamatan paus di wilayah (a) Afrika dan	17
10	Visualisasi <i>zonal statistics</i> antara poligon dan piksel raster	18
11	Box plot nilai MCC dari proses <i>k-fold cross-validation</i> pada model yang dibuat menggunakan (a) seluruh data dan (b) data yang sudah dilakukan <i>undersampling</i>	20
12	(a) Kontribusi variabel berdasarkan <i>gain importance</i> (skala log) dan (b) nilai rata-rata MCC dari proses <i>k-fold cross-validation</i>	22
13	Matriks korelasi linier Pearson pada lima variabel terpilih	23
14	Grafik (a) nilai rata-rata skor SHAP dan (b) <i>gain importance</i> untuk kelima variabel yang digunakan pada model terbaik	24
15	Diagram pencar antara nilai pada setiap variabel terhadap nilai SHAP dengan skala warna yang menunjukkan nilai pada pasangan variabel	25
16	Prediksi zona kemunculan paus pada tanggal 30 November 2023	25

DAFTAR LAMPIRAN

1	Lampiran 1 Penelitian terdahulu mengenai deteksi kemunculan paus	35
2	Lampiran 2 <i>Hyperparameter</i> yang digunakan pada proses pemodelan	36
3	Lampiran 3 <i>Confusion matrix</i> hasil klasifikasi pada data uji	37