

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia, sebuah negara kepulauan raksasa dengan lebih dari 17.000 pulau dan garis pantai yang membentang sekitar 95.000 kilometer, adalah rumah bagi salah satu keanekaragaman hayati ikan terkaya di dunia. Kekayaan ini tidak hanya terbatas pada spesies laut yang hidup di samudra luas, tetapi juga mencakup ribuan jenis ikan air tawar yang mendiami sungai, danau, dan rawa di seluruh nusantara. Posisi geografis yang unik ini menjadikan Indonesia sebagai pusat gravitasi ekologi dan ekonomi di sektor perikanan global. Potensi ekonomi yang signifikan dari sektor perikanan terlihat dari kontribusinya yang substansial terhadap produk domestik bruto nasional dan perannya sebagai sumber mata pencarian bagi jutaan masyarakat pesisir. Lebih dari itu, keanekaragaman spesies ikan ini memiliki nilai ekologis yang tak ternilai, berperan penting dalam menjaga keseimbangan ekosistem akuatik dan rantai makanan global. Namun, di balik potensi yang begitu besar ini, Indonesia juga menghadapi tantangan fundamental, khususnya dalam upaya identifikasi dan pendataan hasil tangkapan nelayan secara akurat dan efisien. Akurasi data ini adalah fondasi esensial untuk pengelolaan sumber daya yang lestari.jiwa.

Indonesia, dengan kekayaan hayati laut yang luar biasa, menghadapi tantangan besar dalam mengelola sektor perikanan tangkap. Masalah utamanya adalah identifikasi spesies ikan yang tidak akurat, yang saat ini masih mengandalkan metode manual. Metode ini tidak efisien, memakan waktu, dan sangat rentan terhadap kesalahan, terutama karena banyaknya spesies yang memiliki kemiripan fisik. Keterlambatan dan ketidakakuratan data ini berdampak langsung pada kebijakan perikanan, berpotensi menyebabkan eksploitasi berlebih pada spesies rentan atau pembatasan yang tidak perlu pada spesies yang melimpah. Konsekuensinya, keberlanjutan populasi ikan, keseimbangan ekosistem, dan mata pencarian nelayan terancam. Oleh karena itu, pengembangan sistem identifikasi ikan yang cepat, akurat, dan mudah diterapkan di lapangan menjadi sebuah keharusan. Integrasi teknologi modern seperti Machine Learning dan Artificial

Intelligence menawarkan solusi inovatif untuk menyediakan data yang andal dan real-time, yang sangat vital untuk pengambilan keputusan yang tepat demi memastikan keberlanjutan sumber daya perikanan Indonesia



Gambar 1. 1 Kekayaan Laut Melimpah tetapi Hadapi Banyak Tantangan [2]

1.2 Perumusan Masalah

1.2.1 Permasalahan

Bagaimana cara membangun model RFC untuk mengklasifikasikan spesies ikan berdasarkan data tangkapan?

Bagaimana kinerja model RFC dalam mengklasifikasikan spesies ikan, diukur dengan metrik akurasi, presisi, recall, dan F1-score?

Bagaimana dampak ketidakseimbangan kelas (class imbalance) pada data tangkapan terhadap performa klasifikasi model RFC?

Fitur-fitur apa saja dari data tangkapan yang paling berkontribusi dalam klasifikasi spesies ikan menggunakan RFC?

Sebagai contoh, penelitian oleh algoritma Random Forest untuk klasifikasi daerah penangkapan ikan dengan akurasi hampir sempurna ($> 99\%$) berdasarkan parameter lingkungan seperti klorofil, suhu, dan salinitas. Namun penelitian tersebut fokus pada geolokasi tangkapan, bukan klasifikasi spesies.

Berbagai metode machine learning telah dikembangkan untuk klasifikasi makhluk hidup, termasuk ikan dan tumbuhan laut. Sebagai contoh menerapkan algoritma Random Forest untuk klasifikasi daerah penangkapan ikan dengan

akurasi tinggi (>99%) berdasarkan parameter lingkungan seperti klorofil, suhu, dan salinitas. Namun, penelitian tersebut berfokus pada geolokasi tangkapan, bukan klasifikasi spesies.

1.2.2 Metode Usulan

Metode usulan dalam penelitian ini bertujuan untuk mengembangkan dan mengimplementasikan model klasifikasi spesies ikan yang berbasis pada algoritma Random Forest Classifier (RFC). Pendekatan ini dirancang untuk mencapai akurasi tinggi dalam identifikasi spesies ikan dari data tangkapan nelayan, yang merupakan prasyarat krusial bagi pengelolaan sumber daya perikanan yang berkelanjutan. Prosedur penelitian ini terbagi menjadi empat tahapan utama yang sistematis yaitu;

1. Pengumpulan dan Pra-pemrosesan Data

Tahap ini berfokus pada persiapan data tangkapan nelayan agar siap untuk pemodelan.

- Pengumpulan Data: Data diambil dari catatan nelayan, mencakup variabel-variabel berikut:

- o Berat Ikan: Menjelaskan karakteristik ukuran ikan (misalnya, 1.5 kg).
- o Lokasi Penangkapan: Menggunakan koordinat geografis (lintang dan bujur) untuk menangkap distribusi spasial spesies (misalnya, -5.000 lintang dan 110.500 bujur).
- o Waktu Penangkapan: Mencakup bulan dan hari dalam minggu untuk mengidentifikasi pola musiman atau harian (misalnya, September, hari Selasa).
- o Spesies Ikan: Variabel target yang menjadi label kelas (misalnya, "Ikan Tuna").

- Pra-pemrosesan Data: Langkah-langkah untuk memastikan kualitas data:

- o Penanganan Nilai Hilang: Mengisi atau menghapus data yang tidak lengkap (misalnya, mengisi 10% data berat yang hilang dengan rata-rata berat ikan sejenis).
- o Penanganan Outlier: Mengidentifikasi dan mengelola nilai ekstrem (misalnya, memeriksa data ikan 1000 kg yang tidak wajar).
- o Encoding Variabel Kategorikal: Mengubah data teks menjadi format numerik (misalnya, "September" menjadi 9).
- o Normalisasi/Standardisasi Data (Opsional): Menyesuaikan skala fitur jika diperlukan.

2. Pembagian Data

Data yang sudah diproses dibagi menjadi dua subset untuk pelatihan dan pengujian model.

- Set Pelatihan (Training Set): Sekitar 70-80% data digunakan untuk melatih model. Model belajar dari pola hubungan antara fitur dan spesies ikan.
- Set Pengujian (Testing Set): Sekitar 20-30% data disimpan untuk mengevaluasi kinerja model. Set ini berisi data yang belum pernah dilihat model, sehingga evaluasi lebih objektif.

3. Pembangunan dan Pelatihan Model RFC

Tahap ini fokus pada pembuatan dan pelatihan model RFC.

- Arsitektur RFC: RFC adalah model ensemble learning yang membangun banyak pohon keputusan secara acak. Metode ini menggunakan bagging (pelatihan pada subset data acak) dan feature randomness (pemilihan fitur acak di setiap node) untuk mengurangi overfitting dan meningkatkan akurasi.

- Konfigurasi Parameter: Kinerja model dioptimalkan dengan menyesuaikan parameter seperti jumlah pohon (n_estimators) dan kedalaman maksimum pohon (max_depth).
- Pelatihan Model: Model dilatih menggunakan set pelatihan, menghubungkan fitur-fitur (berat, lokasi, waktu) dengan spesies ikan.

4. Evaluasi Kinerja Model

Setelah dilatih, model dievaluasi secara menyeluruh menggunakan set pengujian dan metrik-metrik berikut:

- Akurasi: Mengukur persentase prediksi yang benar. Contoh, akurasi 81.12% berarti 81 dari 100 prediksi benar.
- Presisi: Mengukur seberapa banyak prediksi positif yang benar-benar positif. Contoh, presisi 0.8070 berarti 80.70% prediksi spesies tertentu adalah benar.
- Recall (Sensitivitas): Mengukur seberapa banyak kasus positif yang berhasil ditemukan. Contoh, recall 0.8112 berarti model berhasil menemukan 81.12% dari seluruh spesies yang ada.
- F1-Score: Rata-rata harmonik dari presisi dan recall. Contoh, F1-score 0.8072 menunjukkan keseimbangan yang baik antara presisi dan recall.

1.2.3 Pertanyaan Penelitian

1. Berdasarkan permasalahan yang telah diuraikan mengenai tantangan identifikasi spesies ikan yang akurat dan efisien dari data tangkapan nelayan di Indonesia, serta mempertimbangkan metode usulan yang melibatkan algoritma Random Forest Classifier (RFC), penelitian ini merumuskan pertanyaan penelitian utama sebagai berikut:

2. Bagaimana model Random Forest Classifier dapat secara efektif mengklasifikasikan spesies ikan berdasarkan variabel data tangkapan (berat,

lintang, bujur, dan waktu penangkapan) guna meningkatkan akurasi identifikasi spesies, serta bagaimana potensi implementasinya dalam mendukung pengelolaan sumber daya perikanan yang berkelanjutan di Indonesia?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

1. Penelitian Mengembangkan dan mengevaluasi model Random Forest Classifier (RFC) untuk klasifikasi spesies ikan. Model ini akan dibangun berdasarkan data tangkapan nyata yang mencakup variabel berat ikan (dalam kilogram), lokasi penangkapan (koordinat lintang dan bujur spesifik seperti -5.000, 110.500), serta waktu penangkapan (misalnya, bulan September dan hari Selasa).
2. Menganalisis kinerja model RFC dalam mengidentifikasi spesies ikan secara kuantitatif. Analisis ini akan dilakukan dengan mengukur metrik akurasi (misalnya, 81,12%), presisi (misalnya, 0,8070), recall (misalnya, 0,8112), dan F1-score (misalnya, 0,8072), yang semuanya akan diperoleh dari hasil eksperimen.

1.3.2 Manfaat Penelitian

Penelitian Penelitian ini memberikan manfaat langsung bagi pihak-pihak terkait dalam sektor perikanan:

1. Sektor Perikanan Tangkap:

Menyediakan metode identifikasi spesies ikan yang lebih cepat dan akurat daripada identifikasi manual, sehingga meningkatkan efisiensi pendataan dan mendukung pengambilan keputusan berbasis data. Hal ini dapat mengurangi biaya operasional dan mendukung penentuan kuota penangkapan yang lebih baik.

2. Bagi Pengelola Sumber Daya Perikanan dan Pemerintah: Menyediakan data tangkapan yang valid dan terpercaya untuk merumuskan kebijakan yang efektif. Hasil penelitian ini dapat membantu dalam pemantauan stok ikan secara presisi dan meningkatkan efektivitas program konservasi.

3. Masyarakat Umum: Secara tidak langsung, berkontribusi pada kelestarian sumber daya perikanan nasional, yang menjamin ketersediaan ikan sebagai sumber protein dan keberlanjutan mata pencarian nelayan.

4. Manfaat Teoritis

Penelitian ini juga memberikan kontribusi signifikan dalam bidang akademis:

Peneliti dan Akademisi: Menambah literatur di bidang pembelajaran mesin (Machine Learning) dalam konteks biologi kelautan dan ekologi perikanan. Penelitian ini juga menyediakan dasar metodologis untuk pengembangan model klasifikasi yang lebih canggih di masa depan.

1.4 Batasan Masalah

Penelitian ini memiliki beberapa batasan yang perlu diperhatikan dalam menafsirkan hasilnya:

1. Variabel Data Terbatas: Penelitian ini hanya menggunakan berat ikan, lokasi penangkapan (lintang dan bujur), dan waktu penangkapan (bulan dan hari). Data visual (ciri morfologi) atau data dari sensor lingkungan laut (seperti suhu dan salinitas) tidak disertakan.

2. Keterbatasan Sumber Data: Data berasal dari catatan nelayan, sehingga kualitas dan akurasi data sangat bergantung pada ketelitian pencatatan manual. Potensi bias dan kesalahan dalam pengumpulan data tidak dapat dihindari sepenuhnya.

3. Algoritma Klasifikasi Tunggal: Penelitian ini hanya berfokus pada penggunaan algoritma Random Forest Classifier (RFC) tanpa perbandingan komprehensif dengan algoritma pembelajaran mesin lainnya.

4. Lingkup Spesies Ikan Terbatas: Model ini hanya mampu mengklasifikasikan spesies yang terwakili dalam dataset. Kemampuan model untuk mengidentifikasi spesies langka atau yang tidak ada dalam data pelatihan akan terbatas.

5. Faktor Lingkungan dan Musiman Tidak Lengkap: Meskipun waktu penangkapan dipertimbangkan, model mungkin tidak sepenuhnya menangkap variasi distribusi spesies yang dipengaruhi oleh kondisi oseanografi atau pola migrasi ikan yang lebih kompleks.

6. Faktor Operasional dan Eksternal Tidak Termasuk: Penelitian ini tidak mempertimbangkan faktor operasional perikanan seperti jenis alat tangkap,

ukuran kapal, atau keahlian nelayan yang dapat memengaruhi komposisi spesies hasil tangkapan

1.5 Sistematika Proposal

BAB I Pendahuluan :

Bab ini berisi latar belakang penelitian mengenai tingginya penyakit diabetes dan pentingnya deteksi dini menggunakan pendekatan berbasis data. Diuraikan pula permasalahan yang dimiliki data dalam klasifikasi medis serta kebutuhan akan algoritma yang tepat. Bab ini mencakup rumusan masalah, metode usulan, pertanyaan penelitian, tujuan dan manfaat, serta batasan masalah.

BAB II Tinjauan Pustaka :

Bab ini akan menjelaskan teori-teori terkait klasifikasi data dan pengelolaan sumber daya perikanan, dengan fokus pada konsep dasar pembelajaran mesin (khususnya algoritma Random Forest Classifier), pentingnya akurasi data tangkapan untuk keberlanjutan perikanan, serta relevansi penggunaan data spasial dan temporal dalam identifikasi spesies ikan.

BAB III Metodologi Penelitian :

Bab ini menjelaskan metodologi yang digunakan dalam penelitian, meliputi tahapan-tahapan mulai dari pemilihan dataset, preprocessing, pelatihan model menggunakan RFC hingga evaluasi kinerja model menggunakan metrik klasifikasi.

BAB IV Hasil dan Pembahasan :

Pada bab ini akan menguraikan teori dan konsep fundamental yang mendasari penelitian Anda. Pembahasan akan mencakup prinsip pengelolaan perikanan berkelanjutan dan peran akurasi identifikasi spesies, dasar-dasar pembelajaran mesin dengan fokus pada algoritma Random Forest Classifier (RFC), serta relevansi variabel spasial dan temporal dalam klasifikasi spesies ikan.

BAB V Penutup

Bab ini memuat kesimpulan dari hasil penelitian yang menjawab pertanyaan penelitian, serta saran untuk pengembangan penelitian selanjutnya.