

# BAB I PENDAHULUAN

## 1.1. Latar Belakang

Kajian ilmu kelautan sangat penting bagi kelangsungan hidup manusia. Oleh karena itu, perlu dilakukan konservasi sumber daya laut secara berkelanjutan dengan menguasai data dan informasi kelautan yang akan menjadi landasan kognisi dan tata kelola (Zhang, dkk, 2021). Namun, data dan informasi kelautan yang ada belum dikelola dan juga belum difungsikan untuk memprediksi sumber daya laut dengan menggunakan teknik *Machine Learning* (ML). Sehingga penguasaan data dan informasi kelautan menjadi permasalahan dalam konservasi sumber daya laut, karena saat itu masih banyak masyarakat yang memanfaatkan sumber daya laut sebatas memenuhi kebutuhan hidup.

Penelitian ilmiah kelautan telah memasuki era baru *Artificial Intelligence* (AI) yang secara terus-menerus menyempurnakan pengelolaan data kelautan (Lin, 2020). AI seperti ML dapat secara efektif memanfaatkan potensi informasi yang terkandung dalam data kelautan dengan jumlah besar. Oleh karena itu, prospek penerapan metode algoritma ML dalam penelitian ilmiah kelautan cukup menjanjikan, khususnya untuk memantau keanekaragaman hayati laut, memodelkan data *Conductivity, Temperature, Depth* (CTD), dan untuk memprediksi sumber daya laut (Hafeez, dkk, 2019). Selain itu, algoritma ML saat ini banyak digunakan oleh para peneliti dalam mengolah data kelautan dan membangun model prediktif (Bahari, dkk, 2023). Hal lain yang menarik, ML dapat memberikan informasi multi parameter dalam jumlah besar untuk memantau ekosistem laut yang kompleks (Jiang & Zhu, 2022).

Perkembangan keilmuan ML untuk membangun model prediktif saat ini telah banyak digunakan oleh para peneliti, namun spesifikasi penelitian yang berkaitan dengan penggunaan ML untuk memprediksi potensi sumber daya laut merupakan bidang penelitian yang relatif baru. Dengan demikian

penelitian ini diharapkan akan memberi kontribusi data dan informasi mengenai perairan laut Indonesia yang memiliki potensi sumber daya laut.

Adapun harapan kedepannya, data dan informasi kelautan yang memiliki potensi adanya sumber daya laut, akan menjadi referensi sebagai materi penelitian lanjutan yang inovatif bagi para peneliti untuk menguraikan fenomena dan spesifik tertentu di lautan, serta menjadi bahan keputusan investasi bagi para investor maupun pengguna sistem informasi kelautan. Kontribusi lainnya, data dan informasi kelautan menjadi pedoman bagi para nelayan untuk penangkapan ikan yang produktif dengan adanya area subur tempat habitatnya ikan di perairan laut tertentu.

Penggunaan ML untuk memprediksi potensi sumber daya laut membutuhkan data yang akan menjadi variabel prediktor dan diantara data tersebut tentunya harus saling berkorelasi. Data yang dibutuhkan itu adalah data oseanografi Indonesia yang diperoleh dari hasil akuisisi data kelautan oleh para peneliti BPPT yang berkolaborasi dengan para peneliti asing dari berbagai negara, seperti; Amerika, Eropa, Jepang, dan China dengan menggunakan wahana Kapal Riset Baruna Jaya milik BPPT yang berlayar di berbagai perairan laut Indonesia. Dengan demikian pihak BPPT yang poseanografi sebagai bagian dari data kelautan.

Oseanografi merupakan salah satu cabang ilmu kelautan yang sangat luas pokok bahasannya. Namun untuk membatasinya supaya pokok bahasan penelitian lebih fokus pada tujuannya, maka penelitiannya diarahkan pada oseanografi bidang ekologi kelautan yang digali dengan ilmu data sebagai bagian dari ilmu sistem informasi. Adapun tujuannya, untuk menggali informasi mengenai potensi sumber daya laut di perairan laut Indonesia. Metodologi yang akan diterapkan adalah penggunaan data CTD yang akan digunakan sebagai parameter oseanografi untuk dianalisis hubungan korelasinya antara faktor data CTD dengan data salinitas sebagai data target yang menjadi indikator potensi sumber daya laut. Sedangkan potensi sumber daya laut, merupakan hasil analisis prediksi.

Penelitian ini dipicu oleh faktor hasil penelitian terdahulu yang mengungkap, bahwa faktor oseanografi memegang peranan penting terhadap sumber daya laut (Apriliani, dkk, 2021). Sehingga gagasan yang terinspirasi adalah penggunaan teknik ML untuk memprediksi potensi sumber daya laut merupakan bidang penelitian yang relatif baru, serta dapat secara efektif mengeksplorasi potensi informasi yang terkandung dalam data kelautan dengan jumlah besar. Hal tersebut didukung dengan informasi oseanografi tentang data; suhu, kedalaman, dan salinitas dapat memberikan informasi mengenai potensi sumber daya laut (Wright, dkk, 2016). Demikian juga data konduktivitas, dapat memberi informasi mengenai potensi sumber daya laut (Muller, dkk, 2012). Sedangkan faktor salinitas dapat mempengaruhi produksi, distribusi dan umur sumber daya laut (Grilli, dkk, 2020).

Penelitian terdahulu lainnya yang menginspirasi, yaitu penelitian yang menjelaskan mengenai analisis hubungan antara faktor oseanografi dan hasil tangkapan ikan tuna di Laut Banda (Tangke, dkk, 2011). Penelitian terdahulu masih menjadi bagian penting sebagai referensi pada penelitian ini, terutama dalam hal; pada tingkat kedalaman berapa, suhu permukaan laut berapa °C, dan tingkat salinitas berapa, hasil tangkapan ikan dalam jumlah besar akan menjadi representasi atau indikator tingkat kesuburan pada area perairan laut. Hal itu memiliki arti penting dalam merumuskan pola pikir analisis dalam penelitian ini sebagai data ilmiah aktual yang akurat dari hasil penelitian terdahulu yang tidak usang karena waktu. Data hasil penelitian ilmiah terdahulu bersifat tetap dari waktu ke waktu yang terkait dengan siklus bulan dan musim, serta berlaku ukurannya untuk berbagai perairan laut.

Adapun hubungan korelasi antara penelitian terdahulu (Tangke, dkk, 2011) dengan hasil penelitian ini, bahwa nilai salinitas antara 32,4 ppt - 35,1 ppt yang didukung dengan suhu permukaan laut antara 29,1 °C – 29,5 °C dan tingkat kedalaman laut berkisar 0 meter - 400 meter pada suatu perairan laut di Indonesia memiliki tingkat kesuburan sumber daya laut, seperti; perikanan, plankton, dan klorofil-a (Apriliani, dkk, 2021).

Data CTD dan data salinitas akan diproses dengan teknik ML untuk menghasilkan nilai prediksi. Proses akuisisi antara data CTD dan salinitas berbeda karena penggunaan alat survey untuk mengakuisisi data-data tersebut juga berbeda, sehingga tingkat korelasi antara data CTD dan data salinitas harus dibuktikan dengan teknik ML agar diperoleh fakta empiris. Diantara data CTD itu sendiri merupakan data yang saling mempengaruhi, dan data salinitas dipengaruhi oleh data CTD (Ullman & Hebert, 2014). Korelasi yang baik diantara data, maka akan ditunjukkan dengan nilai koefisien korelasi ( $r$ ) +1 atau -1 dengan skala 0 - 1 (Samudrala, 2018) dan akan menjadi indikator dalam memprediksi potensi sumber daya laut. Prosesnya, korelasi antar variabel prediktor yang memiliki pengaruh terhadap variabel target akan menghasilkan nilai prediksi.

Data CTD akan dikelompokkan menjadi variabel prediktor, sedangkan data salinitas dikelompokkan menjadi variabel target. Variabel prediktor disebut juga dengan variabel bebas (*independent*) dan digunakan sebagai data latih pada ML dimaksudkan untuk proses belajar pada mesin. Variabel target disebut juga dengan variabel tidak bebas (*dependent*) dan digunakan sebagai data target yang akan diprediksi oleh data prediktor, sehingga menghasilkan nilai prediksi dalam bentuk bilangan atau numerik. Data latih dimodelkan menggunakan algoritma; *Decision Tree* (DT), *Linear Regression* (LR), dan *Random Forest* (RF), sehingga akan menghasilkan model data. Model itu diuji dengan data uji, hingga menghasilkan model dengan skor terbaik, selanjutnya difungsikan untuk memprediksi data target.

Mengevaluasi akurasi model adalah bagian penting dari proses dalam membuat model ML untuk menggambarkan seberapa baik kinerja model dan prediksinya, tujuannya untuk mendapat model prediksi yang akurat dengan metrik evaluasi. Metrik evaluasi adalah suatu cara untuk memeriksa keakuratan model regresi, karena apabila terdapat kesalahan dalam pemodelan maka akan menjadikan kesalahan pada prediksinya. Metrik evaluasi yang pada umumnya digunakan dalam analisis regresi, yaitu:

- ✓ MAE (*Mean Absolute Error*), yaitu perbedaan antara nilai sebenarnya dan nilai prediksi yang diekstraksi dengan rata-rata perbedaan absolut atas kumpulan data.
- ✓ MSE (*Mean Squared Error*), yaitu perbedaan antara nilai sebenarnya dan nilai prediksi yang diekstraksi dengan kuadrat perbedaan rata-rata atas kumpulan data.
- ✓ RMSE (*Root Mean Square Error*), yaitu tingkat kesalahan oleh akar kuadrat MSE.
- ✓ R2 (*Koefisien Determinasi*), yaitu seberapa baik nilai koefisien dibandingkan dengan nilai yang sebenarnya. Ukuran nilai dari 0 – 1 yang ditafsirkan sebagai presentasi.

MAE, MSE, RMSE pada umumnya berfungsi mengukur nilai kesalahan terkecil untuk prediksi, sedangkan kuadrat koefisien determinasi (R2) berfungsi mengukur selisih antara nilai prediksi dan nilai sebenarnya.

Pada awal sebelum algoritma digunakan untuk pemodelan, diperlukan keputusan untuk memilih algoritma yang tepat dalam menyelesaikan permasalahan regresi pada penelitian data oseanografi CTD. Oleh karena itu, karakteristik datanya harus diteliti terlebih dahulu dengan cara; distribusi data untuk mendapat sebaran data yang baik, ranking data untuk mendapat signifikansi data, dan korelasi data untuk mendapat hubungan antar variabel yang kuat. Sedangkan untuk menentukan kelayakan data oseanografi CTD yang digunakan pada penelitian, hubungan baik diantara data tersebut terlebih dahulu harus diteliti dengan cara menggunakan *scatter plot* yang terdapat pada salah satu alat ukur (*widget*) dalam penggunaan *software orange3* sebagai alat pendukung ML. Penggunaan teknik ML mulai dari pengelompokan data latih dan data uji hingga pemodelan untuk prediksi, didukung dengan *software orange3* dengan tujuan agar mempermudah dalam proses pengolahan dan analisis.

Karakteristik data oseanografi CTD perlu dipahami masing-masing fungsinya pada perairan laut. Konduktivitas berfungsi; menyediakan sumber daya laut, membantu mengatur suhu permukaan laut, membantu mengatur

arus laut, dan membantu menjaga kesehatan ekosistem laut. Suhu berfungsi; mengatur proses kehidupan dan persebaran organisme, serta mempengaruhi jumlah oksigen ( $O_2$ ) yang terlarut dalam air. Semakin rendah suhu permukaan laut, maka semakin besar kelarutan oksigen di dalam air, begitu pula sebaliknya. Suhu permukaan laut di perairan laut Indonesia umumnya berkisar antara  $28\text{ }^{\circ}\text{C}$  -  $31\text{ }^{\circ}\text{C}$ . Kedalaman; secara umum terbagi menjadi dua yaitu perairan laut dangkal dan perairan laut dalam. Perairan laut dangkal merupakan zona laut mulai dari garis sudut terendah hingga kedalaman sekitar 120 - 200 meter, selebihnya merupakan kategori laut dalam. Faktor kedalaman laut berhubungan erat dengan suhu vertikal akibat penetrasi cahaya. Salinitas merupakan bagian dari parameter oseanografi, namun merupakan parameter terpisah dari data CTD (Mensah, dkk, 2009). Salinitas air laut diukur dalam satuan ppt, sedangkan air laut normal memiliki rata-rata 35 ppt yang setara dengan 35 gram garam terlarut per kilogram air laut (Hutton & Roy, 2023).

## 1.2. Tujuan Penelitian

Tujuan penelitian ini adalah: *pertama*, membuat model data oseanografi Indonesia dengan menggunakan metode ML; *kedua*, memprediksi adanya potensi sumber daya laut dari model data yang telah dibuat.

## 1.3. Manfaat Penelitian

Manfaat dari penelitian ini adalah: *pertama*, memberi kontribusi data dan informasi perairan laut Indonesia yang berpotensi adanya sumber daya laut yang diharapkan menjadi referensi sebagai materi penelitian lanjutan yang inovatif bagi para peneliti untuk menguraikan fenomena dan spesifikasi tertentu di lautan; *kedua*, memberi kontribusi data dan informasi untuk bahan keputusan investasi bagi para investor maupun pengguna sistem informasi kelautan serta menjadi pedoman bagi para nelayan untuk penangkapan ikan yang produktif dengan adanya area subur tempat habitatnya ikan di perairan laut tertentu di Indonesia.