

PREDIKSI KUALITAS AIR BERDASARKAN KELARUTAN AIR MENGGUNAKAN METODE *MACHINE LEARNING* PADA BUDIDAYA SALMON

Rizty Maulida Badri¹

¹Magister Sains Komputasi Institut Teknologi Bandung, Jawa Barat, Indonesia

Abstrak. Budidaya perikanan darat (*land-based aquaculture*) akhir-akhir ini diminati di Indonesia. saat ini ITERA memiliki salmon center untuk mengembangkan ikan salmon jenis *Rainbow Trout (Oncorhynchus mykiss)*. Namun, Kualitas air sangat penting dalam budidaya ikan karena kualitas air yang buruk dapat mempengaruhi kesehatan dan pertumbuhan ikan. Meskipun telah banyak ditemukan aplikasinya di bidang hidrologi, namun penerapan *Machine Learning* pada prediksi kualitas air pada budidaya salmon belum dapat penulis temukan. Untuk itu, penulis akan menggunakan *Random Forest Clasifier*, *Light Gradient- Boosted machine (LightGBM)* dan *Decision Tree Clasifier* dalam penelitian ini. Dari penelitian ini, model prediksi kualitas air yang paling akurat adalah *random forest clasifier* yang tanpa di setel *hyperparameter tuning* akurasi nya paling tinggi, koefisien antar fitur sangat rendah, dan berdasarkan grafik dari variabel yang ada, 39 % air dapat diminum. , kualitas air ini juga tidak cukup baik untuk budidaya salmon

Kata Kunci. Kualitas air, *Machine Learning*, Salmon

PENDAHULUAN

Budidaya perikanan darat (*land-based aquaculture*) akhir-akhir ini diminati di Indonesia. Salmonid adalah keluarga ikan salmon yang sudah banyak dikenal di Indonesia. Ikan jenis ini memiliki kandungan protein yang tinggi sehingga dapat memenuhi kebutuhan protein di Indonesia. Karena hal itu , saat ini ITERA memiliki salmon center untuk mengembangkan ikan salmon jenis *Rainbow Trout (Oncorhynchus mykiss)*.

Akan tetapi, dengan meningkatnya produksi budidaya, resiko penyakit menular dapat meningkat dengan cepat sehingga pengelolaan parasit merupakan tantangan bagi budidaya ikan bersirip[1]. Salah satu parasit yang ada dalam peternakan salmon adalah Kutu laut. Kutu laut adalah *copepoda ektoparasit* yang merumput di kulit dan lendir salmon yang mengakibatkan lesi, masalah osmoregulasi, dan infeksi sekunder yang dapat menyebabkan masalah yang parah baik

pada salmon liar maupun salmon yang dibudidayakan[2].

Kualitas air sangat penting dalam budidaya ikan karena kualitas air yang buruk dapat mempengaruhi kesehatan dan pertumbuhan ikan. Kualitas air ditentukan oleh faktor fisik-kimiawi dan biologis yang secara langsung atau tidak langsung mempengaruhi kelangsungan hidup, pertumbuhan dan reproduksi ikan. Dua faktor eksternal terpenting yang mengarahkan metabolisme ikan adalah suhu [3] dan ketersediaan oksigen [4]. Suhu menentukan laju metabolisme, dan dengan demikian laju pertumbuhan, sedangkan O_2 terlarut adalah faktor pembatas utama untuk metabolisme aerobik[5]. Pengamatan perilaku kelompok salmon telah mencatat perubahan distribusi vertikal sebagai respons terhadap berbagai stimulus lingkungan termasuk cahaya, suhu, salinitas, pemberian makan, kecepatan arus air, suara dan tingkat infestasi kutu laut. Namun, penghindaran kondisi O_2 terlarut yang sub optimal tidak konsisten), dan dalam lingkungan yang heterogen dapat digantikan oleh faktor lain[6].

Dalam beberapa tahun terakhir, banyak peneliti telah membangun model untuk memprediksi kualitas air. Model-model ini dapat dibagi menjadi dua

kategori: model prediksi kualitas air teoretis dan model prediksi kualitas air berbasis data[7]. Dengan pesatnya perkembangan kecerdasan buatan (AI), *deep learning* telah menjadi salah satu teknologi yang paling populer untuk penelitian prediksi deret waktu hidrologi dalam beberapa tahun terakhir karena pemetaan nonlinier yang kuat dan kemampuan belajar, toleransi kesalahan yang tinggi, dan kemampuan generalisasi yang lebih baik[8]. Misalnya, Ma et al. (2020) menggunakan teknologi *deep learning* untuk memprediksi nilai BOD_5 air pelabuhan perkotaan. Metode ini tidak hanya dapat memprediksi nilai BOD_5 secara akurat, tetapi juga memecahkan masalah matriks sparse secara lebih cerdas dengan hasil yang memuaskan. Yu dkk. (2020) menggunakan model long-short-term memory (LSTM) yang dikombinasikan dengan *denoising domain wavelet* untuk mensimulasikan perubahan historis konsentrasi klorofil *a* di danau dan memprediksi tren masa depannya. Hasil mereka menunjukkan bahwa model ini memiliki kesalahan prediksi yang lebih rendah dan kemampuan generalisasi yang lebih tinggi. Zhang dkk. (2018) mengumpulkan data dari struktur saluran pembuangan gabungan (CSO) yang dipantau oleh Internet of Things dan membuat model *deep learning* yang berbeda untuk mensimulasikan dan

memprediksi ketinggian air dari struktur CSO. Hasil mereka menunjukkan bahwa metode LSTM dan gated recurrent unit (GRU) menunjukkan kemampuan yang kuat dalam prediksi deret waktu multi-langkah. Selain itu, kurva belajar untuk GRU lebih cepat.

Meskipun telah banyak ditemukan aplikasinya di bidang hidrologi, namun penerapan *Machine Learning* pada prediksi kualitas air pada budidaya salmon belum dapat penulis temukan. Hal ini membuat penulis sangat tertantang untuk memprediksi kualitas air Berdasarkan Kelarutan Air menggunakan metode *machine learning* pada budidaya Salmon, mengingat kualitas air sangat berpengaruh dalam produksi ikan salmon. Untuk itu, penulis akan menggunakan *Random Forest Classifier*, *Light Gradient- Boosted machine (LightGBM)* dan *Decision Tree Classifier* dalam penelitian ini.

METODOLOGI

Dataset didapatkan dari *kaggle* dan merupakan data yang di dapatkan dari sungai di india. Terdapat 3276 data dan 10 variabel dalam dataset tersebut. Setelah dilakukan *data cleaning* selanjutnya dilakukan visualisasi data untuk melihat korelasi diantara 10 variabel tersebut.

Selanjutnya dilakukan data *preprocessing* dimana 20% untuk test data dan 80% untuk eksperimen. Penulis menggunakan tiga metode klasifikasi dalam penelitian ini, yaitu *Random Forest Classifier*, *Light Gradient- Boosted machine (LightGBM)* dan *Decision Tree Classifier*

Random Forest Classifier adalah metode pohon gabungan pengembangan metode Classification and Regression Tree (CART) dengan menerapkan metode bootstrap aggregating (bagging) dan random feature selection. Random forest merupakan analisis nonparametrik yang dapat menghasilkan akurasi yang lebih tinggi, dapat mengatasi data latih dalam jumlah yang besar secara efisien, dan tidak terdapat pemangkasan peubah seperti pada algoritma decision tree (pohon klasifikasi tunggal)[9].

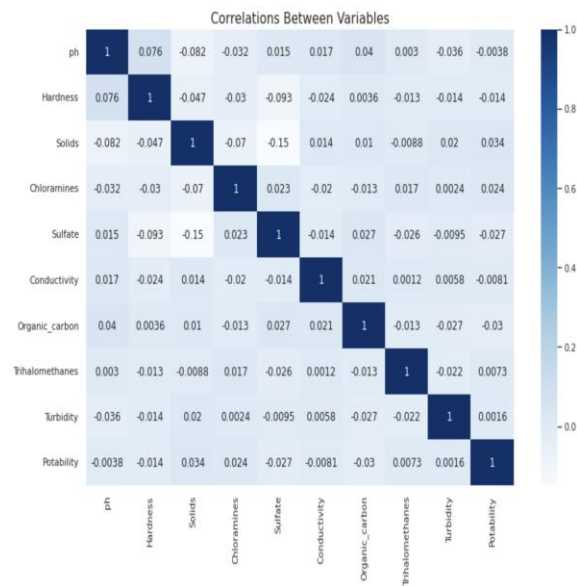
Light Gradient- Boosted machine (LightGBM) adalah kerangka kerja peningkatan gradien berdasarkan pohon keputusan untuk meningkatkan efisiensi model dan mengurangi penggunaan memori. Model ini menggunakan dua teknik baru: **Pengambilan Sampel Satu Sisi Berbasis Gradien** dan **Paket Fitur Eksklusif (EFB)** yang memenuhi batasan algoritma berbasis histogram yang terutama digunakan di semua kerangka kerja GBDT (Gradient Boosting Decision Tree)[10].

Decision Tree Classifier adalah algoritma pembelajaran Mesin yang diawasi. Ini digunakan dalam algoritma klasifikasi dan regresi. Pohon keputusan seperti pohon dengan node. Cabang tergantung pada sejumlah faktor. Ini membagi data menjadi cabang-cabang seperti ini sampai mencapai

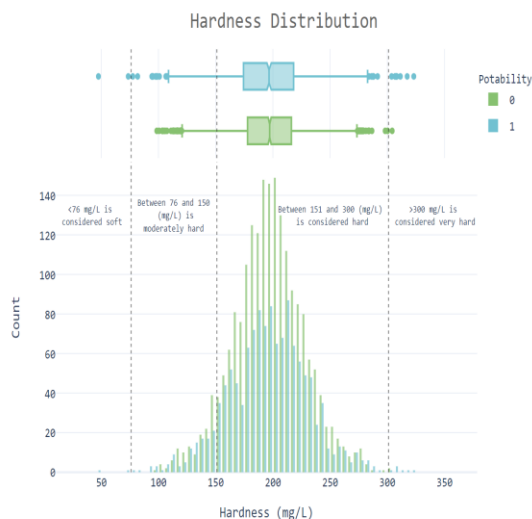
nilai ambang batas. Sebuah pohon keputusan terdiri dari simpul akar, simpul anak, dan simpul daun[9].

HASIL DAN PEMBAHASAN

Pada Gambar 1. Ditunjukkan korelasi antar variabel. Dapat dilihat bahwa korelasi antar variabel sangat rendah.

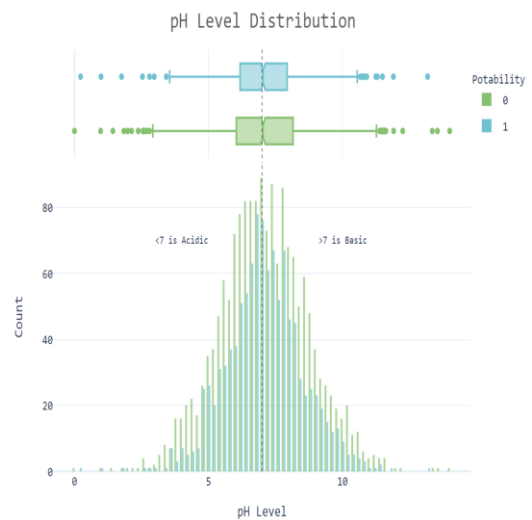


Gambar 1. Korelasi antar variabel



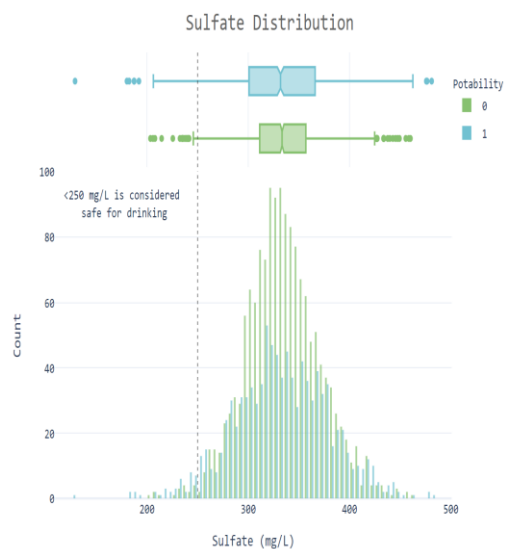
Gambar 2. Grafik tingkat kesadahan air

Dari gambar 2 diatas dapat dilihat bahwa berdasarkan tingkat kesadahan air, air yang dapat diminum lebih dari 90%.



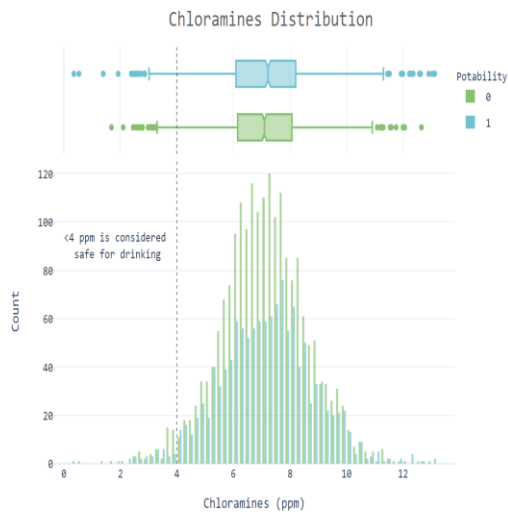
Gambar 2. Grafik distribusi level Ph

Dari gambar 3 diatas dapat dilihat bahwa berdasarkan distribusi level pH, jumlah asam dan basa yang terkandung dalam air seimbang, sehingga air layak untuk diminum.



Gambar 4. Grafik distribusi tingkat sulfat

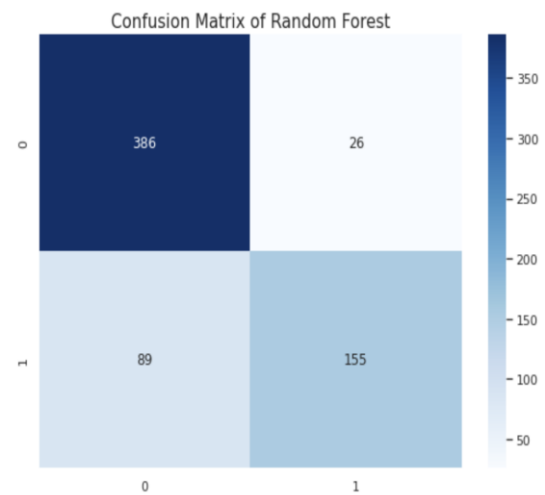
Dari gambar 4 diatas dapat dilihat bahwa berdasarkan distribusi tingkat sulfat, hanya 1,8% air yang aman untuk diminum.



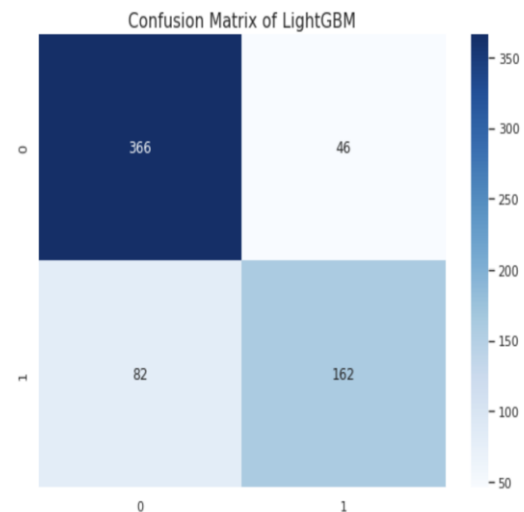
Gambar 5. Grafik distribusi tingkat Chloramines

Dari gambar 5 diatas dapat dilihat bahwa berdasarkan distribusi tingkat kloramin, hanya 2% air yang aman untuk diminum.

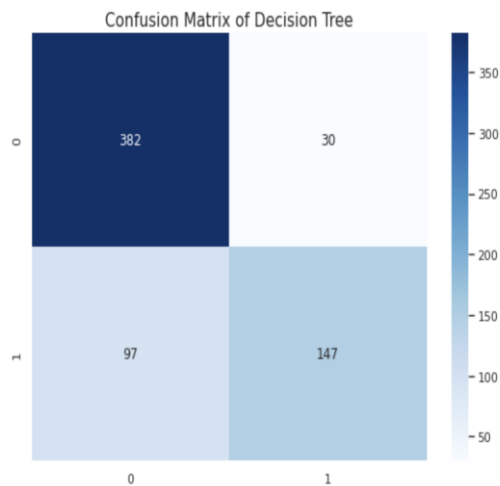
Confusion Matriks adalah pengukuran performa untuk masalah klasifikasi *machine learning* dimana keluaran dapat berupa dua kelas atau lebih. *Confusion Matrix* adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. Ada empat istilah yang merupakan representasi hasil proses klasifikasi pada *confusion matrix* yaitu True Positif, True Negatif, False Positif, dan False Negatif [11].



Gambar 6. Confusion matrix random forest clasifier



Gambar 7. Confusion matrix LightGBM



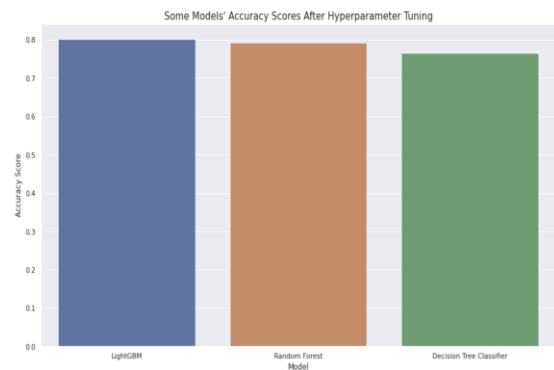
Gambar 8. Confusion matrix Decision Tree Clasifier

Dari Gambar 6,7, dan 8 diatas terlihat bahwa *decision tree clasifier* yang prediksi benarnya paling tinggi dan *random forest classifier* yang prediksi salahnya paling tinggi.

Tabel 1. Tingkat Akurasi dari model yang diuji sebelum disetel.

Nomor	Model	Akuraasi
1	Random Forest	0,824695
2	LightGBM	0,804878
3	Decision Tree	0,746951

Dari Tabel 1 diatas terlihat bahwa *random forest clasifier* memiliki nilai akurasi tertinggi.



Gambar 9. Nilai akurasi dari model yang diuji setelah Hyperparameter tuning.

Tabel 2. Tingkat Akurasi dari model yang diuji setelah disetel.

Nomor	Model	Akuraasi
1	Random Forest	0,792683
2	LightGBM	0,801829
3	Decision Tree	0,765244

Dari Tabel 2 diatas terlihat bahwa *LightGBM* memiliki nilai akurasi tertinggi.

KESIMPULAN

Kesimpulan dari penelitian ini adalah bahwa model prediksi kualitas air yang paling akurat adalah *random forest clasifier* tanpa di setel hyperparameter tuning akurasi nya paling tinggi, koefisien antar fitur sangat rendah, dan berdasarkan grafik dari variabel yang ada, 39 % air dapat diminum. dan berdasarkan variabel kelarutan air yang diuji, kualitas air ini juga tidak cukup baik untuk budidaya salmon karena nilai – nilai variabel kelarutan airnya terlalu tinggi untuk kolam salmon. Hal ini bisa menyebabkan kematian pada salmon.

DAFTAR PUSTAKA

1. Murray AG, Peeler EJ (2005) A framework for understanding the potential for emerging diseases in aquaculture. *Prev Vet Med* 67: 223–235.
2. Pike AW, Wadsworth SL (2000) Sealice on salmonids: their biology and control. *Adv Parasitol* 44:233–337
3. E. Eliason, A. Farrell (2016) **Oxygen uptake in Pacific salmon *Oncorhynchus* spp.: when ecology and physiology meet**
4. *Journal of Fish Biology*, 88 (2016), pp. 359-388
5. J.G. Richards, A.P. Farrell, C.J. Brauner (2009) **Hypoxia** Academic Press
6. Oppedal F, Juell JE, Johansson D. (2007). Thermo- and photoregulatory swimming behaviour of caged Atlantic salmon: implications for photoperiod management and fish welfare. *Aquaculture* 265: 70–81
7. G. Claireaux, D. Chabot (2016) **Responses by fishes to environmental hypoxia: integration through Fry's concept of aerobic metabolic scope** *Journal of Fish Biology*, 88 (2016), pp. 232-251
8. Oppedal F, Taranger GL, Juell JE, Fosseidengen JE, Hansen T. (1997). Light intensity affects growth and sexual maturation of Atlantic salmon (*Salmo salar*) postsmolts in sea cages. *Aquat Living Resour* 10: 351–357.
9. Ramadhan, Aditya Susetyo, Budi Indahwati.(2019). **Pemodelan** Klasifikasi Random Forest untuk Mengidentifikasi Faktor Penting dalam Meningkatkan Mutu Pendidikan. URL : <https://repository.ipb.ac.id/handle/123456789/100010>
10. <https://papers.nips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html>
11. Jianfeng Xu, Yuanjian Zhang, Duogian Miao.(2020).Three-way confusion matrix for classification: A measure driven view. *Information Sciences Volume 507*. Elsevier