**PREDIKSI SALINITAS AIR LAUT DENGAN**

**DEEP NEURAL NETWORK**

**Wiwien Widyastuti 1, J. B. Budi Darmawan 2**

**Dosen Universitas Sanata Dharma 1,2**

**Paingan, Maguwoharjo, Depok, Sleman**

**Sur-el : wiwien@usd.ac.id 1,** **b.darmawan@usd.ac.id** **2**

***Abstract :*** *Research in the marine field is important to look at the state of the ocean's atmosphere and the biodiversity that can live in it. Based on the CalCOFI data set, seawater salinity correlates with the depth of the sea. Sea water salinity can be predicted based on its depth. The classic method that is often used is the least squares regression. Deep neural network is one of the machine learning methods that has been widely applied to regression problems. This study aims to find better predictive performance by comparing the least squares regression method and the deep neural network method. The research method is done first by making an equation model with the least squares regression method. Second, by training the deep neural network using the same data so that the network model is obtained. Furthermore, the results of both methods are compared by calculating MAE and MSE. The results showed that the network model with superior deep neural network was used to predict data outside the range of trained data compared to the least squares regression method.*

***Keywords:*** *Deep Neural Network, CalCOFI, regression*

*Abstrak : Penelitian di bidang kelautan penting dilakukan untuk melihat keadaan atmosfir laut dan keanekaragaman hayati yang dapat hidup di dalamnya. Berdasarkan data set CalCOFI, salinitas air laut berkorelasi dengan kedalaman laut. Salinitas air laut dapat diprediksi berdasarkan kedalamannya. Metode klasik yang sering digunakan adalah regresi kuadrat terkecil. Deep neural network merupakan salah satu metode machine learning yang telah banyak diterapkan untuk masalah regresi. Penelitian ini bertujuan untuk mencari unjuk kerja prediksi yang lebih baik dengan membandingkan metode regresi kuadrat terkecil dan metode deep neural network. Metode penelitian dilakukan pertama dengan membuat model persamaan dengan metode regresi kuadrat terkecil. Kedua, dengan melatih jaringan deep neural network menggunakan data yang sama sehingga diperoleh model jaringan. Selanjutnya hasil dari kedua metode tersebut dibandingkan dengan menghitung MAE dan MSE. Hasil penelitian menunjukkan bahwa model jaringan dengan deep neural network lebih unggul digunakan untuk memprediksi data di luar kisaran data yang dilatih dibandingkan dengan metode regresi kuadrat terkecil.*

***Kata kunci****: Deep Neural Network, CalCOFI, regresi*

1. PENDAHULUAN

Kumpulan data CalCOFI merupakan data oseanografi dan larva ikan terpanjang (1949-sekarang) dan terlengkap di dunia [1]. Data fisik, kimia, dan biologis yang dikumpulkan pada interval waktu dan daerah tertentu merupakan data yang berharga untuk mendokumentasikan siklus iklim di Arus California dan serangkaian respon biologis terhadapnya. [1]. Salah satu data CalCOFI adalah data tentang salinitas dan kedalaman. Penelitian oseanografi terbatas karena data yang dikumpulkan merupakan cuplikan sehingga menghasilkan data diskret. Pemodelan diperlukan untuk keperluan prediksi atau ekstrapolasi.

Saat ini teknologi *machine learning* berperan besar dalam berbagai aspek kehidupan masyarakat modern. Salah satu aplikasi *machine learning* yang sudah banyak diteliti digunakan untuk prediksi, dengan hasil yang memuaskan. *Deep neural network* merupakan metode/ algoritma yang saat ini banyak digunakan pada teknologi *machine learning*.

Penggunaan teknologi *machine learning* untuk prediksi telah banyak dilakukan oleh para peneliti. Farifteh, Van der Meer, Atzberger,dan Carranza menggunakan regresi kuadrat terkecil parsial dan jaringan syaraf tiruan (ANN) untuk memprediksi salinitas tanah [2]. Weicheng Wu, et al menggunakan regresi *machine learning* untuk prediksi salinitas tanah di Central Mesopotamia, Irak [3]. Hari, Rammani, dan Sharan menggunakan model *feed-forward deep neural network* untuk prediksi harga persediaan pasar [4]. Semua penelitian yang telah dilakukan menunjukkan bahwa *machine learning* memberikan hasil yang baik.

Saat ini, Graphics Processing Units(GPU) mempunyai unjuk kerja yang lebih baik daripada CPU dalam operasi *floating point* dan lebar pita memori. Salah satu *platform computing* dan model pemrograman dalam GPU adalah Compute Unified Device Architecture (CUDA). GPU dengan CUDA menawarkan unjuk kerja yang tinggi dengan biaya yang sangat murah , dan dapat diintegrasikan ke dalam sistem komputer berunjuk kerja tinggi [5][6]. Pemrograman paralel menggunakan GPU dengan CUDA ini juga telah terbukti memperbaiki kecepatan komputasi untuk menyelesaikan permasalahan *elasticity* [7].

Kemampuan pemrosesan paralel menggunakan GPU dengan CUDA ini sangat mendukung bidang Artificial Intelligence khususnya dalam bidang teknologi *machine learning*. Pemrograman *Python* sering digunakan untuk mengembangkan model jaringan *deep neural network*. Pustaka yang digunakan adalah *tensorflow* dan *keras* yang digunakan untuk pembuatan model jaringan *deep neural network* dan dapat memanfaatkan GPU untuk pemrosesan paralel sehingga waktu yang diperlukan untuk pelatihan menjadi lebih singkat.

Penelitian ini bertujuan untuk mencari unjuk kerja prediksi yang lebih baik dengan membandingkan unjuk kerja prediksi salinitas air laut dengan dua model, yaitu model regresi kuadrat terkecil dan model *deep Neural Network* (NN) dengan lapisan yang *fully-connected*.

1. **METODOLOGI PENELITIAN**

Metode penelitian dilakukan dengan tahap:

1. Menentukan data set
2. Membuat model dengan regresi kuadrat terkecil.
3. Membuat model dengan *deep NN*
4. Menganalisa hasil
	1. Menentukan Data Set

Data set yang digunakan adalah data set yang diambil dari data set CalCOFI tentang salinitas dan kedalaman air laut dari satu lokasi yang sama. Salinitas diukur dengan banyaknya kandungan garam dalam air laut (satuan g/kg, 1 g garam dalam 1 kg air laut). Lokasi ini mempunyai data salinitas dan kedalaman air laut dari 1 sampai 4108m. Data pada lokasi ini sebanyak 54 data pada kedalaman 1 sampai 3000 digunakan untuk pelatihan atau pembuatan model baik dengan metode regresi kuadrat terkecil maupun dengan metode *deep NN*. Data selebihnya digunakan untuk pengujian model.

* 1. **Regresi Kuadrat Terkecil**

Data seringkali diberikan untuk nilai-nilai diskrit sepanjang suatu rangkaian kesatuan. Namun kadang-kadang kita perlu menaksir nilai pada titik-titik di antara nilai diskrit tersebut. Teknik-teknik untuk mencocokkan kurva terhadap data-data yang demikian diperlukan agar memperoleh taksiran-taksiran antara. Salah satu teknik pencocokan kurva jika data cuplikan masih mengandung galat adalah regresi kuadrat terkecil. Strategi regresi adalah menurunkan suatu kurva tunggal yang mewakili kecenderungan umum dari data. Kurva dirancang untuk mengikuti pola titik-titik sebagai suatu kelompok. Teknik regresi kuadrat terkecil akan menurunkan suatu kurva yang meminimumkan ketidaksesuaian antara titik-titik data dan kurva [8].

Fungsi polinomial derajat ke-m secara umum berbentuk persamaan (1)

 (1)

*a0, a1, …, am* merupakan koefisien yang dapat dihitung dengan metode penyelesaian persamaan linear simultan :

* 1. **Membuat Model dengan Regresi Kuadrat Terkecil**

Data akan diolah dengan regresi kuadrat terkecil linear (orde 1) sehingga memperoleh persamaan berbentuk (2) berikut :

 (2)

Variabel y adalah nilai prediksi untuk salinitas. Variabel x merupakan kedalaman air laut (dalam meter). Koefisien a+ merupakan letak perpotongan garis dengan sumbu y dan koefisien a1 merupakan kemiringan garis regresi. Koefisien a0 dan a1 dihitung dengan rumus (3) dan (4)

 (3)

 (4)

Dengan :



* 1. **Deep Neural Network**

Lapisan

input

Lapisan

Tersembunyi 1

Lapisan

Tersembunyi 2

Lapisan

output

X­1

Z­6

Z­1

Z­2

Z­3

Z­4

Z­5

Z­7

Q­6

Q­1

Q­2

Q­3

Q­4

Q­5

Q­7

Y­1

…..

..

**Gambar 1. Arsitektur *Deep NN***

*Deep neural network* (*Deep NN*) merupakan pengembangan dari *artificial neural network* (ANN) konvensional. *Deep NN* menggunakan lebih dari satu lapisan tersembunyi.. Penelitian ini menggunakan *deep NN* yang terkoneksi penuh menggunakan 2 buah lapisan tersembunyi. Arsitektur jaringannya dapat dilihat pada gambar 1.

Unit input disimbolkan dengan *Xi*, dengan *i*=1,…,*n*. Unit tersembunyi 1 disimbolkan dengan *Zj*, dengan *j*=1,..,*m*. Unit tersembunyi 2 disimbolkan dengan *Qk*, dengan *k*=1,…,*p*. Unit output disimbolkan dengan *Yl*, dengan *l*=1,…,*r*. Bobot yang menghubungkan unit input dengan lapisan tersembunyi 1 disimbolkan dengan *vij*. Bobot yang menghubungkan unit tersembunyi 1 dan unit tersembunyi 2 disimbolkan dengan *sjk*. Bobot yang menghubungkan unit tersembunyi 2 dan unit output disimbolkan dengan *wpl*.

* 1. **Membuat Model dengan Deep NN**

Untuk mendapatkan model yang terbaik dengan *error* yang minimum untuk keluaran salinitas air laut, berbagai struktur ANN dengan jumlah lapisan tersembunyi yang berbeda (dari 1 sampai 3) , jumlah neuron/unit yang berbeda untuk setiap lapisan tersembunyi (dari 1 sampai 10) digunakan untuk pelatihan dan pengujian.

Arsitektur *deep neural network* terdiri dari 1 unit masukan, lapisan tersembunyi, dan 1 unit keluaran. Semua unit pada semua lapisan terkoneksi penuh (*fully connected*) dengan unit-unit pada lapisan yang berdekatan. Arsitektur jaringan dapat dilihat pada gambar 1. Setelah menentukan arsitektur jaringan, tahap selanjutnya adalah pelatihan jaringan dan pengujian. Pelatihan jaringan menggunakan algoritma *backpropagation*. Fungsi aktivasinya menggunakan ReLu (*Rectified Linear Unit*). Dari hasil percobaan diperoleh model ANN yang terbaik menggunakan 2 lapisan tersembunyi masing-masing dengan 7 unit tersembunyi. Parameter jaringan yang digunakan : *optimize*=*adam* ; *learning rate* = 0,001 ; *loss function* = MSE; *epoch*=180.

Pelatihan jaringan dilakukan dengan menggunakan perangkat lunak *Python* 3.6.8 berbasis Windows 7, dengan pustaka *Tensorflow* 1.13.1 dan *Keras* 2.3.4. Perangkat keras yang digunakan adalah sebuah komputer dengan spesifikasi prosesor Intel Core 2 Quad Q6600, RAM 8 GB, dilengkapi GPU NVIDIA GT 730 384 core memory 2GB dengan CUDA.

* 1. **Menganalisa Hasil**

Pengujian dianalisa dengan menghitung MAE dan MSE dari hasil prediksi masing-masing model dengan data sebenarnya. Model diuji menggunakan data uji yang bervariasi. Kemudian dibandingkan nilai MAE dan MSE dari kedua model tersebut.

Dalam penelitian ini dilakukan pengamatan menggunakan dua skenario yang akan dijelaskan di bawah ini. Untuk kedua skenario ini, data untuk pelatihan menggunakan data salinitas air laut yang sama sejumlah 54 data dari kedalaman 1 sampai sekitar 3000 meter.

* 1. **MAE dan MSE**

*Mean Absolute Error* (MAE) dan *Mean Squared Error* (MSE) adalah metode untuk mengukur tingkat keakuratan suatu model regresi. Nilai MAE merepresentasikan rata – rata kesalahan *(error)* absolut antara hasil peramalan dengan nilai sebenarnya.

MAE dirumuskan dengan persamaan (5)

 (5)

Dengan *yi* adalah nilai-nilai peramalan, *xi* adalah nilai-nilai sebenarnya, *n* adalah banyaknya data, *ei* adalah *error* atau selisih antara nilai peramalan dan nilai sebenarnya.

Sedangkan nilai MSE adalah nilai rata-rata dari kuadrat kesalahan(*error*) atau selisih antara hasil peramalan dengan nilai sebenarnya.

MSE dirumuskan dengan persamaan (6)

 (6)

1. HASIL DAN PEMBAHASAN

Skenario pertama membandingkan unjuk kerja kedua model dengan data salinitas air laut untuk pelatihan dan pengujian menggunakan 54 data dari kedalaman 1 sampai sekitar 3000 meter dengan data pelatihan dan pengujian dari lokasi yang sama. Grafik data target dan data prediksi skenario 1 dapat dilihat pada gambar 2.



**Gambar 2. Grafik skenario 1**

Skenario kedua membandingkan unjuk kerja kedua model dengan data salinitas air laut untuk pelatihan menggunakan 54 data dari kedalaman 1 sampai 3000 meter dan pengujian menggunakan 66 data dari kedalaman 1 sampai sekitar 4000 meter dengan data pelatihan dan pengujian dari lokasi yang sama. Grafik data target dan data prediksi skenario 2 dapat dilihat pada gambar 3.



**Gambar 3. Grafik skenario 2**

Hasil pengamatan MAE dan MSE untuk keempat scenario menggunakan model regresi orde 1 dan model *deep neural network* di atas disajikan dalam tabel 1.

**Tabel 1. Nilai MAE dan MSE untuk kedua skenario**

|  |  |  |
| --- | --- | --- |
|  | Skenario 1 | Skenario 2 |
| MAE regresi | 0,287397136 | 0,364300297 |
| MAE *deep NN* | 0,036098882 | 0,052356594 |
| MSE regresi | 0,107038849 | 0,18400637 |
| MSE *deep NN* | 0,004238042 | 0,006135054 |

Dari skenario 2 pada tabel 1, untuk data pelatihan dengan kedalaman 1 sampai sekitar 3000 meter dan data pengujian dengan kedalaman 1 sampai 4108 meter di lokasi yang sama dengan data pelatihan dengan model *deep NN* diperoleh MAE sebesar 0,052356594 sedangkan menggunakan model regresi orde 1 diperoleh MAE sebesar 0,364300297. Hal ini menunjukkan bahwa *deep NN* dapat digunakan untuk memprediksi salinitas air laut berdasarkan kedalaman air laut dengan lebih baik dibanding dengan regresi kuadrat terkecil orde 1 dengan data pengujian di dalam kisaran data pelatihan di lokasi yang sama.

Dari tabel 1 terlihat dari skenario 1 ke skenario 2 terjadi peningkatan MAE regresi sebesar 0,076903161 dan peningkatan MAE *deep NN* sebesar 0,016257712 sedangkan peningkatan MSE regresi sebesar 0,076967521 dan peningkatan MSE *deep NN* sebesar 0,001897013. Semakin kecil nilai MAE dan MSE maka semakin baik prediksinya. Dalam skenario 1 dan skenario 2 ini menggunakan data pelatihan salinitas air laut yang sama sejumlah 54 data dari kedalaman 1 sampai 3000 meter untuk lokasi yang sama, perbedaannya dalam skenario 1 data pengujian salinitas air laut sama dengan data pelatihan yaitu sejumlah 54 data dari kedalaman 1 sampai 3000 meter untuk lokasi yang sama sedangkan dalam skenario 2 data pengujian salinitas air laut menggunakan sejumlah 66 data dari kedalaman 1 sampai 4108 meter untuk lokasi yang sama Peningkatan MAE dan MSE model regresi lebih besar dari peningkatan MAE dan MSE model *deep NN* ini menunjukkan model *deep NN* lebih baik daripada model regresi orde 1 dalam memprediksi data pengujian di luar kisaran data pelatihan. Dalam hal ini data pelatihan dari kedalaman 1 sampai 3000 meter sedang data pengujian dari kedalaman 1 sampai 4108 meter di lokasi yang sama.

1. KESIMPULAN

Kesimpulan yang dapat diambil dari penelitian ini bahwa; *Deep NN* dapat digunakan untuk memprediksi salinitas air laut berdasarkan kedalaman air laut dengan lebih baik dibanding dengan regresi kuadrat terkecil orde 1 dengan data pengujian di dalam kisaran data pelatihan. Selain itu unjuk kerja *deep NN* lebih baik dibanding dengan regresi kuadrat terkecil orde 1 dalam memprediksi salinitas air laut berdasarkan kedalaman air laut dengan data pengujian di luar kisaran data pelatihan di lokasi yang sama.

**DAFTAR PUSTAKA**

1. Sohier, “[California Cooperative Oceanic Fisheries Investigations](http://calcofi.org/about-calcofi.html) (CalCOFI) data set,”DiaksesMei 2019. <https://www.kaggle.com/sohier/calcofi>
2. Farifteh, J., van der Meer, F., Atzberger, C., & Carranza, E, ”Quantitative analysis of salt affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN),” *Remote Sensing of Environment*, 110, 59–78, 2007
3. Weicheng Wu, et al, “Soil salinity prediction and mapping by machine learning regression in Central Mesopotamia, Iraq,” [Online] Available: *wileyonlinelibrary.com/journal/ldr. Land Degrad Dev.* 2018;29:4005–4014, 2018
4. Hari K.C., Rammani Adhikari 2, Er. Sharan Thapa, “Feed-Forward Deep Learning Model for Data Analysis and Prediction,” *International Journal of Computer Trends and Technology (IJCTT) –* Volume 63 Number 1. 2018
5. Coates A, Huval B, Wang T, Wu D J and Ng A Y., “Deep learning with COTS HPC Systems,” *Proceedings of the International Conference on Machine Learning (ICML) (Atlanta, GA, USA) 1337,* 2013
6. Tan Y, Ding K, “ A Survey on GPU-based Implementation of Swarm Intelligence Algorithms,” *IEEE Transactions on Cybernetics* 46 2028, 2016.
7. Darmawan J. B. B. and Mungkasi S, “Parallel Computations Using a Cluster of Workstations to Simulate Elasticity Problems,” *Journal of Physics: Conference Series accepted*. 2016
8. Steven C. Chapra, Raymond P, Canale, *“Numerical Methods for Engineers*,”6th edition. McGraw Hill. New York.2010