
IMPLEMENTASI METODE K-NEAREST NEIGHBOUR UNTUK MENDUGA SALINITAS AIR LAUT

Yuri Efenie

Fakultas Teknik, Program Studi Sistem Informasi

Universitas Islam Madura

Email: yuri.efenie.2016@gmail.com

Miftahul Walid

Fakultas Teknik, Program Studi Teknik Informatika

Universitas Islam Madura

Email: miftahul.walid@uim.ac.id

ABSTRAK

Dalam penelitian mencoba untuk menduga salinitas air laut menggunakan metode K- Nearest Neighbour, metode ini berfungsi untuk mengklasifikasi data input dengan menggunakan metode pengukuran jarak (Distance) dengan data training, variable yang digunakan dalam penelitian ini adalah nilai letak kordinat (latitude dan longitude) dan outputnya adalah berupa nilai salinitas, studi kasus dalam penelitian ini adalah perairan selatan sumenep, Sistem yang dibuat telah mampu melakukan pendugaan namun dengan tingkat error yaitu rata – rata 1,00 sehingga perlu adanya analisa kembali karena data yang digunakan hanya sedikit, perlu adanya penambahan data sehingga hasil akan lebih optimal, perlu juga melakukan percobaan dengan mengubah metode atau menyederhanakan aturan – aturan atau dengan menambahkan variabel input dalam sistem yang telah dibuat sehingga menghasilkan nilai akurasi yang lebih baik, karena sistem yang ada masih membutuhkan waktu yang cukup lama dalam melakukan pendugaan.

Kata kunci: K-NN, Salinitas, Latitude, Longitude

ABSTRACT

In this research, trying to predict the salinity of sea water using the K-Nearest Neighbor method, this method serves to clarify the input data using the distance measurement method with training data, the variable used in this study is the value of the location of coordinates (latitude and longitude) and the output is in the form of salinity, the case study in this study is the southern waters of Sumenep, the system has been able to make an estimate but with an error rate of 1.00 so that there is a need for re-analysis because the data used is only small, the need for additional data so that the results will be more optimal, it is also necessary to experiment with changing methods or simplifying rules or by adding input variables in the system that have been created so that it produces better accuracy values, because the existing system still requires a long time in estimating.

Keywords: K-NN, Salinity, Latitude, Longitude

1. PENDAHULUAN

Perkembangan teknologi informasi pada saat ini telah masuk ke berbagai bidang ilmu, seperti pendidikan, pertanian, perikanan, kelautan dan lain lain, pada beberapa bulan yang lalu pertanian garam nasional mendapatkan permasalahan karena habisnya stok garam sehingga terjadi lonjakan harga garam dan pemerintah harus melakukan kebijakan impor garam dari luar negeri, hal ini merupakan fenomena yang kurang baik karena notabeni Indonesia merupakan negara yang memiliki garis pantai terpanjang kedua di dunia dengan hal itu indonesia sangat berpotensi dalam pengembangan pertanian garam. salinitas merupakan kadar garam yang terlarut dalam air, semakin tinggi salinitas maka semakin banyak garam yang

akan dihasilkan, walaupun kadar salinitas baik masih belum tentu menentukan kualitas garam yang baik, dalam penelitian ini peneliti hanya ingin melakukan pendugaan salinitas air laut, beberapa penelitian untuk menduga salinitas air laut sudah pernah dilakukan oleh Wouthuyzen pada tahun 2011 dalam penelitian ini peneliti mengukur salinitas permukaan laut menggunakan data warna laut menggunakan citra modis (Wouthuyzen dkk, 2011), pada tahun 2016 Nafizah melakukan penelitian dengan judul Evaluasi Algoritma Wouthuyzen dan Son untuk Pendugaan Sea Surface Salinity (SSS) (Studi Kasus: Perairan Utara Pamekasan) (Nafizah dkk, 2016), dalam penelitian tersebut menunjukkan bahwa perbedaan nilai salinitas perairan dari algoritma Wouthuyzen dengan data in-situ berkisar antara 2,66 hingga 3,91 psu. Sedangkan selisih salinitas algoritma Son dengan data in-situ 2,14 hingga 3,00 psu. Adapun nilai NMAE algoritma Wouthuyzen dengan data in-situ sebesar 7,2 %, sedangkan pada algoritma Son sebesar 10,6 %, karena masih besarnya tingkat error maka dalam penelitian ini peneliti mencoba menggunakan metode klasifikasi yang bersifat supervised yaitu metode K-Nearest Neighbour (K-NN), metode K-NN sudah sering digunakan untuk pattern recognition, seperti pola tulisan tangan angka Arabic (Akbar & Sarwoko, 2016) pengenalan pola bentuk bunga (Akbar & Sarwoko, 2016) klasifikasi daun dengan perbaikan fitur citra (Liantoni, 2015) pola citra dalam deteksi penyakit kulit (Liyanto, Roro, 2015), K-NN juga pernah digunakan untuk sistem keamanan rumah (Junoh et al., 2012) algoritma K-Nearest Neighbor dalam klasifikasi data hasil produksi kelapa sawit pada PT. Minamas Kecamatan Parindu (Krisandi, Prihandono, & Bayes, 2013) Perpaduan Genetic Algorithm dan K-NN juga pernah digunakan untuk deteksi kanker (Bhuvaneswari & Therese, 2015), Variable yang digunakan dalam penelitian ini adalah nilai letak koordinat (latitude dan longitude) dan outputnya adalah berupa nilai salinitas, studi kasus dalam penelitian ini adalah perairan selatan sumenep, konsep kerja sistem ini adalah dengan cara mengklasifikasi data input menggunakan metode perhitungan jarak dengan data training, diharapkan dengan metode K-NN bisa melakukan pendugaan lebih baik dari metode sebelumnya. Tulisan yang dapat dimuat adalah yang mengkaji masalah yang berhubungan dengan teknologi dibidang mesin, elektronika dan ilmu komputer.

2. STUDI PUSTAKA DAN LANDASAN TEORI

2.1. Penelitian Terdahulu

Febri Liantoni pada tahun 2015 melakukan penelitian untuk mengklasifikasi daun dengan perbaikan fitur citra menggunakan metode K-Nearest Neighbour (K-NN), Dari hasil pengujian yang telah dilakukan meliputi pra-proses, ekstraksi fitur dan klasifikasi didapatkan kesimpulan. Penggunaan median filter dan erosi citra mampu memperbaiki citra daun, sehingga dapat mempermudah proses ekstraksi tur dan klasifikasi. Pengujian klasifikasi dengan menggunakan metode K-Nearest Neighbor (KNN) didapatkan nilai akurasi sebesar 86,67%, hasil ini menunjukkan metode KNN mampu melakukan klasifikasi daun dengan baik (Liantoni, 2015).

Pada tahun 2016 Roni Akbar dan Eko Adi Sarwoko melakukan penelitian yang berjudul studi analisis pengenalan pola tulisan tangan angka *arabic* (INDIAN) menggunakan metode *K-Nearest Neighbors* dan connected component labeling Pada studi analisis ini telah dilakukan pengujian dengan menggunakan 100 citra uji. Tiga hasil klasifikasi terbaik dari pengenalan pola tulisan tangan angka Arabic (Indian) menggunakan metode *k-nearest neighbors* (KNN) adalah 86% saat nilai $k = 1$, 84% dengan nilai $k = 3$, dan 83% dengan $k = 5$ (Akbar & Sarwoko, 2016).

Sebastian Rori Lisyanto pada tahun 2015 melakukan penelitian untuk mengenali pola citra penyakit kulit menggunakan K-NN penelitian ini menggunakan 60 kali uji coba. Uji coba ini dilakukan dengan mencari akurasi klasifikasi dari setiap kemungkinan kombinasi sudut, jarak dan jumlah data training dan testing yang berbeda-beda. Kemudian dicari rata-rata dari semua uji coba berdasarkan sudut, jarak dan jumlah data training dan testing sehingga dapat diketahui pengaruh ketiganya terhadap akurasi klasifikasi. Pada hasil klasifikasi didapatkan akurasi tertinggi dengan menggunakan sudut 45° , jumlah data training dan testing 150 dan 60 serta jarak 3 piksel (Liyanto, Roro, 2015).

Pengenalan pola bentuk bunga menggunakan principle component analysis dan K-NN pernah dilakukan oleh Herfina pada tahun 2013 PCA digunakan untuk proses ekstraksi fitur yaitu untuk mendapat penciri dari citra bunga tersebut. Sedangkan K-NN digunakan untuk proses klasifikasi citra bunga. Hasil evaluasi pengenalan pola pada bentuk citra bunga menggunakan K-NN berdasarkan ekstraksi ciri dengan PCA menghasilkan akurasi sebesar 97,70% (Herfina, 2013).

2.1 K-Nearest Neighbour

Algoritme k-nearest neighbor (K-NN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data training yang jaraknya paling dekat dengan objek tersebut. Pengelompokkan suatu data baru berdasarkan jarak data baru itu ke beberapa data/tetangga (neighbor) terdekat. Dalam hal ini jumlah data/tetangga terdekat ditentukan oleh user yang dinyatakan dengan k. Misalkan ditentukan k=5, maka setiap data testing dihitung jaraknya terhadap data training dan dipilih 5 data training yang jaraknya paling dekat kepada testing. (Herfina, 2013)

Algoritma K-NN adalah suatu metode yang menggunakan algoritma supervised Perbedaan antara supervised learning dengan unsupervised learning adalah pada supervised learning bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru. Sedangkan pada unsupervised learning, data belum memiliki pola apapun, dan tujuan unsupervised learning untuk menemukan pola dalam sebuah data(Krisandi et al., 2013).

Metode k-Nearest Neighbour (K-NN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan ciri dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak Euclidean dengan rumus umum sebagai berikut :

$$d = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \dots \dots \dots (1)$$

Klasifikasi k-NN dilakukan dengan mencari k- buah tetangga terdekat dan memilih kelas dengan ki terbanyak pada kelas ωi. Adapun langkah-langkah klasifikasi k-NN adalah sebagai berikut (Walid & Darmawan, 2017) :

1. Jika sekumpulan vector data latih memiliki N titik data secara keseluruhan, maka kenallah k- buah tetangga terdekat dari sampel x dimana k merupakan bilangan ganjil.
2. Dari k-buah tetangga terdekat tersebut, identifikasikan jumlah vector ki pada kelas ωi, = 1,2,.....,m dengan Σi ki = k. k
3. Masukkan x pada kelas ωi dengan nilai ki maksimum.
4. Jika terdapat dua atau lebih kelas ωi yang memiliki E tetangga terdekat. Maka terjadilah kondisi seimbang dan digunakan strategi pemecahan konflik.
5. Untuk masing-masing kelas yang terlibat dalam konflik, tentukan jarak di antara x dengan kelas ω .
6. Jika pola pelatihan ke-m dari kelas ωi yang terlibat dalam konflik ditunjukkan dengan y^{im} = {y₁^{im}, ..., ..., y_N^{im}} maka jarak antara x dengan kelas adalah :

$$d_i = \frac{1}{E} \sum_{j=1}^n |(x_j - y_j^{im})| \dots \dots \dots (2)$$

7. Masukkan x ke dalam kelas dengan jarak di paling kecil, yaitu x ∈ ω , jika dc < di untuk i, C ∈ 1,....., M dan i ≠ C.

3. METODOLOGI PENELITIAN

3.1 Proses Pengumpulan Data

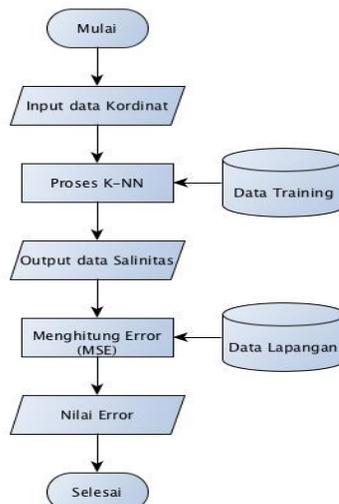
Data diambil di perairan utara kabupaten sumenep, data yang diamati adalah salinitas perairan dengan menggunakan refractometer untuk mengukur salinitas air laut dan GPS untuk mengetahui letak koordinat pengamatan, Pengambilan data lapangan tersebar di 22 titik stasiun pengamatan pada tanggal 10 - 21 Juli 2017. Adapun sebaran stasiun pengamatan dapat dilihat pada tabel berikut.

Tabel 1. Data Lapangan

Stasiun	Lintang	Bujur	Salinitas
1	7,17826	113,9164	31
2	7,18193	113,9085	33
3	7,21143	113,8924	32
4	7,20828	113,912	33
5	7,20684	113,9403	34
6	7,19546	113,8826	33
7	7,12554	113,7964	30
8	7,11524	113,7375	31
9	7,10942	113,6889	32
10	7,13647	113,5922	34
11	7,1037	113,4221	33
12	7,1408	113,3528	35
13	7,1756	113,3022	34
14	7,1446	113,2627	34
15	7,1447	113,2337	35
16	7,1426	113,215	35
17	7,20447	113,9312	34
18	7,17649	113,9198	31
19	7,1027	113,4013	33
20	7,1153	113,3745	34
21	7,1215	113,3642	35
22	7,1528	113,2731	34

3.2 Perancangan Sistem

Adapun perancangan sistem pendugaan salinitas air laut adalah sebagai berikut :



Gambar 1. Perancangan Sistem

Data input berupa data nilai kordinat (*Longitude* dan *Latitude*) kemudian diproses oleh metode K-NN dengan mencari nilai jarak terdekat (*Euclidean distance*) dengan data *training*, setelah beberapa iterasi dan mendapat nilai dengan jarak terdekat, proses K-NN akan selesai dan akan memunculkan nilai salinitas (*output*). Proses selanjutnya mencari nilai *error* dari output hasil perhitungan K-NN dengan menggunakan metode *Means Square Error* (MSE), dengan MSE akan diketahui berapa nilai perbedaan antara hasil K-NN dengan data lapangan (*real*).

4. HASIL

Metode K – Nearest Neighbour membutuhkan proses learning untuk menghasilkan keputusan yang akurat dalam proses Testing, dari 22 data lapangan, 15 data digunakan sebagai data Training dan 7 data digunakan sebagai data Testing, setelah dilakukan proses menggunakan metode K-NN didapatkan hasil sebagai berikut :

Tabel 2. Hasil dari proses K-NN

ST	Long	Lat	Salinitas	Prediksi
7	7,12554	113,7964	30	33
8	7,11524	113,7375	31	33
9	7,10942	113,6889	32	33
10	7,13647	113,5922	34	33
11	7,1037	113,4221	33	33
12	7,1408	113,3528	35	35
13	7,1756	113,3022	34	34

Setelah dilakukan perhitungan dengan menggunakan Mean Square Error (MSE) untuk mengetahui nilai error data prediksi menggunakan metode K-NN yang dibandingkan dengan data salinitas lapangan. Adapun hasil perhitungan adalah sebagai berikut :

Tabel 3. Hasil perhitungan MSE

ST	Salinitas (Lap)	Prediksi (K-NN)	MSE
7	30	33	3
8	31	33	2
9	32	33	1
10	34	33	1
11	33	33	0
12	35	35	0
13	34	34	0

5. KESIMPULAN

Sistem yang dibuat telah mampu melakukan pendugaan namun dengan tingkat error yaitu rata – rata 1,00 sehingga perlu adanya analisa kembali karena data yang digunakan hanya sedikit, perlu adanya penambahan data sehingga hasil akan lebih optimal, perlu juga melakukan percobaan dengan mengubah metode atau menyederhanakan aturan – aturan atau dengan menambahkan variabel input dalam sistem yang telah dibuat sehingga menghasilkan nilai akurasi yang lebih baik, karena sistem yang ada masih membutuhkan waktu yang cukup lama dalam melakukan pendugaan.

DAFTAR PUSTAKA

- [1] Akbar, R., & Sarwoko, E. A. (2016). Studi Analisis Pengenalan Pola Tulisan tangan angka arabic (indian) Menggunakan Metode K-Nearest Neighbours dan Connected Component Labelling. *Dinamika Rekayasa*, 12(2), 45–51.
- [2] Bhuvaneswari, P., & Therese, A. B. (2015). Detection of Cancer in Lung With K-NN Classification Using Genetic Algorithm. *Procedia Materials Science*, 10(Cnt 2014), 433–440. <https://doi.org/10.1016/j.mspro.2015.06.077>
- [3] Herfina. (2013). PENGENALAN POLA BENTUK BUNGA MENGGUNAKAN PRINCIPLE COMPONENT ANALYSIS. In *Seminar Nasional Teknologi Informasi dan multimedia 2013* (pp. 25–30).
- [4] Junoh, A. K., Mansur, M. ., Abu, S. ., Wan Ahmad, W. ., Mukhtar, A., & Fauzy, S. . (2012). Home Security System Based on K-NN Classifier. *Procedia Engineering*, 38, 1811–1815. <https://doi.org/10.1016/j.proeng.2012.06.223>
- [5] Krisandi, N., Prihandono, B., & Bayes, N. (2013). ALGORITMA k - NEAREST NEIGHBOR DALAM KLASIFIKASI DATA HASIL PRODUKSI KELAPA SAWIT PADA PT . MINAMAS. *Buletin Ilmiah Math. Stat. Dan Terapannya (Bimaster)*, 2(1), 33–38.
- [6] Liantoni, F. (2015). Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor. *ULTIMATICS*, VII(2), 98–104. Retrieved from <http://library.umn.ac.id/jurnal/public/uploads/papers/pdf/baafa3456d29008c73a7e2cdb3f613f4.pdf>
- [7] Liyanto, Roro, S. (2015). Implementasi K-Nearest Neighbour untuk mengenali pola citra dalam deteksi penyakit kulit. *Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Dian Nuswantoro Semarang*. Retrieved from www.eprints.dinus.ac.id/15325/1/jurnal_15398.pdf
- [8] Nafizah, Muhammad Jaelani, L., & Winarso, G. (2016). Evaluasi Algoritma Wouthuyzen dan Son untuk Pendugaan Sea Surface Salinity (SSS) (Studi Kasus: Perairan Utara Pamekasan). *Jurnal Teknik ITS*, 5(2).
- [9] Walid, M., & Darmawan, A. K. (2017). Pengenalan Ucapan Menggunakan Metode Linear Predictive Coding (LPC) dan K-nearest Neighbor, 7(1), 13–22.
- [10] Wouthuyzen, S. dkk. (2011). MEASURING SEA SURFACE SALINITY OF THE JAKARTA BAY USING REMOTELY SENSED OF OCEAN COLOR DATA ACQUIRED BY MODIS SENSOR, 36(2), 51–70.