

Perbandingan Algoritma kNN, C4.5, dan Naive Bayes dalam Pengklasifikasian Kesegaran Ikan Menggunakan Media Foto

Wella¹, Ni Made Satvika Iswari², Ranny³

Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara, Tangerang, Indonesia

wella@umn.ac.id

satvika@umn.ac.id

ranny@umn.ac.id

Diterima 26 Oktober 2017

Disetujui 20 Desember 2017

Abstract— Indonesia merupakan negara kepulauan yang memiliki berbagai jenis keanekaragaman ikan. Potensi perikanan laut sebesar 6,5 juta ton per tahun, namun jumlah produksinya hanya mencapai 5,06 juta ton. Hal ini menunjukkan proses produksi belum optimal. Proses produksi serta pemilahan yang masih tradisional membuat produksi berjalan lambat. Dalam penelitian ini dikembangkan sebuah metode untuk mengklasifikasikan kesegaran ikan berdasarkan citra digital ikan. Adapun algoritma yang digunakan adalah kNN, C4.5, dan Naive Bayes. Berdasarkan hasil uji coba yang dilakukan, algoritma kNN memberikan nilai akurasi yang tertinggi diantara algoritma lainnya. Sehingga kNN dinilai cocok digunakan untuk mengklasifikasikan kesegaran ikan. Metode yang dihasilkan dalam penelitian ini diharapkan dapat membantu mengotomatisasi proses produksi yang sebelumnya manual.

Index Terms — kNN, C4.5, Naive Bayes, Pengolahan Citra Digital, Tingkat Kesegaran Ikan.

I. PENDAHULUAN

Indonesia sebagai negara kepulauan memiliki keanekaragaman jenis hasil laut terutama ikan. Produksi ikan menjadi salah satu potensi ekonomi yang dapat diunggulkan pada daerah di Indonesia. Potensi perikanan di Indonesia pada tahun 2010 sebesar 65 juta ton, tapi hasil produksi budidaya ikan hanya sebesar 10,19 juta ton atau sebesar 16 persen[1]. Potensi perikanan laut sebesar 6,5 juta ton per tahun tapi jumlah produksinya hanya 5,06 juta ton[1]. Proses produksi serta pemilahan yang masih tradisional membuat produksi berjalan lambat.

Pada ilmu komputer dan teknologi informasi, telah banyak dikembangkan metode pencarian pola citra digital untuk memecahkan masalah. Pada penelitian ini dicari pola citra digital fisik luar ikan untuk menentukan kesegaran ikan. Pada industri perikanan untuk menentukan tingkat kesegaran ikan terdapat beberapa cara antara lain secara biokimiawi dan secara visual. Secara kimiawi, ikan diukur kadar bakteri,

semakin tinggi jumlah bakteri pada ikan berarti semakin rendah tingkat kesegaran ikan. Pengukuran secara visual dilakukan dengan lebih sederhana yaitu dengan melihat penampakan luar ikan, misal warna kulit, kekenyalan daging, serta warna insang. Pengukuran secara visual lebih sering dilakukan pada proses produksi karena memerlukan waktu yang lebih singkat, namun pengukuran dengan cara ini memiliki beberapa kelemahan. Kelemahan pengukuran dengan cara visual yaitu memerlukan ahli yang telah memiliki kemampuan dalam memilah ikan segar. Diharapkan dengan didukung teknologi pemilahan dengan teknologi citra digital akan mempercepat dan mengurangi kesalahan dalam proses pemilahan.

Metode pengolahan citra digital dapat mencari pola atau ciri tertentu pada gambar, ciri yang didapat digunakan untuk mengidentifikasi kesegaran ikan. Tingkat kesegaran ikan bergantung pada masa/ jangka waktu proses pembusukan ikan dimulai dari saat ikan mati berada di udara terbuka. Semakin lama ikan mati berada di udara terbuka, maka tingkat kesegaran ikan akan semakin menurun. Pada penelitian ini, klasifikasi ikan busuk adalah berdasarkan bau busuk atau tidak segar. Sementara klasifikasi ikan segar adalah berdasarkan bau segar dan wajar.

Hasil dari pencarian pola citra digital ini kemudian diproses lebih lanjut dengan menggunakan Algoritma k-Nearest Neighbor, C4.5, dan Naive Bayes untuk mengklasifikasikan kesegaran. Ketiga algoritma ini dipilih karena termasuk dalam 10 algoritma yang paling banyak digunakan dalam *machine learning*[2]. Hasil perbandingan dari ketiga algoritma ini diharapkan dapat digunakan untuk menentukan algoritma yang memberikan akurasi terbaik. Adapun batasan dalam penelitian ini adalah hanya 3 jenis ikan yang digunakan, yaitu ikan gurame, nila, dan kakap. Kemudian data pola citra digital yang digunakan dibatasi hanya pada bagian mata ikan.

II. TINJAUAN PUSTAKA

A. Pengukuran Tingkat Kesegaran Ikan

Tingkat kesegaran ikan berpengaruh pada proses pembusukan yang terjadi di ruang terbuka tanpa kondisi pendinginan. Semakin lama ikan berada di ruang terbuka semakin rendah tingkat kesegarannya. Pada bidang perikanan terdapat 4 pembagian kelas mutu ikan[3]:

1. Prima (kesegaran ikan sangat baik sekali)
2. Advanced (kesegaran ikan masih baik)
3. Sedang (kesegaran ikan sudah mulai berkurang)
4. Mutu rendah (busuk)

Terdapat parameter yang digunakan untuk mengukur tingkat kesegaran ikan yang terdiri atas beberapa faktor [3][4]:

1. Faktor Fisikawi
2. Faktor Kimiawi
3. Faktor Mikrobiologik
4. Faktor Sensorik

Pada penelitian yang akan dilakukan menggunakan parameter fisikawi karena gambar digital yang digunakan hanya dapat menangkap ciri fisik ikan. Pengukuran parameter kesegaran ikan pada penelitian ini akan menggunakan teknik pengujian pada[4].

B. Pengolahan Citra Digital

Dalam bidang teknik informatika pengolahan citra digital terdiri dari beberapa tahap, antara lain[5]:

1. Pembentukan Citra Digital yaitu menangkap gambar objek sebagai data citra digital menggunakan kamera digital berupa nilai Red, Green, Blue. Hasil data berupa bilangan antara 0 hingga 255. Citra digital akan disimpan pada format citra digital seperti JPEG, BMP, GIF, dan lain sebagainya.
2. Image Preprocessing yaitu mengatur tingkat kontra warna gambar serta menghilangkan gangguan/ noise yang dapat mempengaruhi ketajaman gambar.
3. Image Segmentation dan deteksi tepi yaitu membagi area citra digital antara bagian latar gambar dengan gambar objek, serta memberikan tanda pada tepi objek.
4. Seleksi dan ekstraksi ciri dilakukan pada tahap akhir pemrosesan citra digital dengan tujuan mengambil ciri pada masing-masing objek citra digital sehingga dapat digunakan sebagai pembeda dengan objek lain yang sejenis.

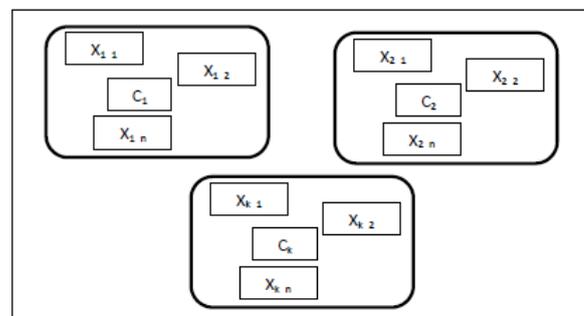
5. Representasi dan deskripsi merupakan pencatatan lokasi dan koordinat ciri pada citra digital.
6. Pengenalan pola dilakukan untuk mengetahui pola gambar yang menjadi ciri suatu objek.
7. Interpretasi citra merupakan pemberian label pada hasil objek yang telah berhasil dikenali.
8. Penyusunan Basis Pengetahuan membuat referensi sebagai data pengenalan objek.

C. K- Nearest Neighbor

Metode pengenalan pola yang memiliki tingkat akurasi tinggi salah satunya adalah k Nearest Neighbor (kNN), sebuah metode yang menghitung jarak kemiripan antar data, data yang memiliki jarak kemiripan terdekat akan dikelompokkan menjadi satu kelompok yang sama. Penentuan jumlah kelompok (k) dapat dilakukan dengan berbagai metode, antara lain ditentukan secara acak, atau ditentukan secara pasti. Pada penelitian ini jumlah kelompok ditentukan berdasarkan jumlah kadar gula yang digunakan pada eksperimen. Pengukuran jarak akan menggunakan jarak Euclidean. Berikut adalah Algoritma kNN yang digunakan pada penelitian ini:

1. Memasukkan nilai jumlah nilai kelas yang ingin dibentuk secara random (k)
2. Memasukkan data nilai yang ingin dikelompokkan (X)
3. Pilih secara acak pusat kelas sebanyak k dari data X (C)
4. Hitung jarak kemiripan tiap anggota X terhadap C
5. Data dengan jarak kemiripan terkecil dari masing-masing C akan dipilih sebagai anggota C.

Sehingga jika digambarkan hasil pengelompokan kNN sesuai dengan Gambar 1.



Gambar 1 Hasil Pengelompokan dengan kNN

Perhitungan jarak antar data menggunakan metode jarak Euclidean. Ciri gambar dari hasil perhitungan FFT akan digunakan sebagai input data di kNN ini. Setiap hasil pemrosesan citra digital akan dilatih untuk memperoleh kelompok data berdasarkan jarak

kemiripannya. Data ciri gambar yang memiliki jarak kemiripan terdekat akan dikelompokkan menjadi satu kelompok.

D. C4.5

C4.5 merupakan salah satu algoritma yang paling banyak digunakan di komunitas pembelajaran *data mining*. Selain itu C4.5 telah menjadi standar komunitas *de facto*[6]. Algoritma C4.5 membuat pohon keputusan (*decision tree*) dengan strategi *divide* dan *conquer*. Di C4.5, setiap simpul di pohon dikaitkan dengan serangkaian kasus. Juga, kasus diberi bobot untuk memperhitungkan nilai atribut yang tidak diketahui[7].

E. Naive Bayes

Naive Bayes merupakan pendekatan probabilistik yang membuat asumsi kuat tentang bagaimana data dihasilkan, dan memberi kesan model probabilistik yang mewujudkan asumsi tersebut. Kemudian menggunakan koleksi contoh pelatihan berlabel untuk memperkirakan parameter model generatif. Klasifikasi pada contoh baru dilakukan dengan peraturan Bayes dengan memilih kelas yang paling mungkin menghasilkan contoh. Klasifikasi Naive Bayes adalah model yang paling sederhana, karena mengasumsikan bahwa semua atribut adalah independen[8].

III. METODE PENELITIAN



Gambar 2 Metode Pengklasifikasian Kesegaran Ikan

Pada penelitian ini dilakukan pengklasifikasian kesegaran ikan dengan metode seperti yang ditunjukkan pada Gambar 2. Tahap pertama yang dilakukan adalah melakukan input citra digital ikan. Ikan dapat difoto pada salah satu sisinya. Tahapan kedua yang dilakukan adalah memotong foto pada bagian mata ikan. Hal ini dilakukan karena hanya warna pada bagian mata saja yang digunakan untuk klasifikasi. Tahapan selanjutnya adalah melakukan peringkasan warna dari foto mata ikan. Adapun acuan nilai RGB yang digunakan ditunjukkan pada Tabel 1,

Tabel 2, Tabel 3. Adapun acuan nilai RGB ini ditentukan berdasarkan observasi umum di lapangan pada saat pengambilan data. Tahapan terakhir adalah pengklasifikasian kesegaran ikan dengan menggunakan 3 algoritma yang telah disebutkan sebelumnya, yaitu kNN, C4.5, dan Naive Bayes.

Tabel 1 Acuan Nilai RGB Foto Ikan Gurame

RGB						
	Rmin	Rmax	Gmin	Gmax	Bmin	Bmax
Hitam	13	46	14	41	2	32
Coklat	33	128	19	95	3	68
Abu- Abu	24	212	19	202	31	184
Kuning	53	139	32	93	6	50

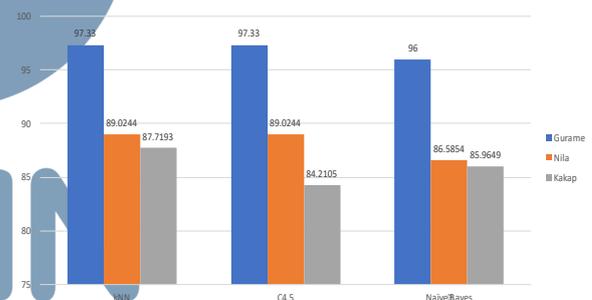
Tabel 2 Acuan Nilai RGB Foto Ikan Nila

RGB						
	Rmin	Rmax	Gmin	Gmax	Bmin	Bmax
Black	2	88	1	69	12	92
Red	54	201	7	155	19	141
Grey	5	211	18	210	26	206
Yellow	102	216	52	178	21	118

Tabel 3 Acuan Nilai RGB Foto Ikan Kakap

RGB						
	Rmin	Rmax	Gmin	Gmax	Bmin	Bmax
Black	13	75	3	69	2	79
Red	33	201	8	174	3	164
Grey	24	222	19	222	31	225
Yellow	53	156	32	149	6	106

IV. HASIL PERBANDINGAN AKURASI ALGORITMA KLASIFIKASI KESEGERAN IKAN



Gambar 3 Perbandingan Akurasi Algoritma Klasifikasi Kesegaran Ikan

Eksperimen dilakukan dengan menggunakan tools WEKA (<https://www.cs.waikato.ac.nz/ml/weka/>). Berdasarkan eksperimen klasifikasi kesegaran ikan yang telah dilakukan dengan menggunakan algoritma kNN, C4.5, dan Naive Bayes, seperti yang ditunjukkan pada Gambar 3, dapat dilihat bahwa Algoritma kNN memberikan akurasi paling tinggi diantara Algoritma C4.5 dan Naive Bayes.

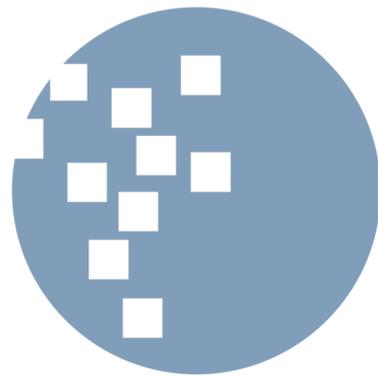
Berdasarkan hasil yang telah diperoleh, Algoritma kNN dinilai cocok untuk kasus pengklasifikasian kesegaran ikan berdasarkan citra digital ikan. Adapun hal ini dikarenakan kNN memiliki beberapa kelebihan, yaitu tangguh terhadap data latih yang banyak memiliki *noise* dan efektif untuk data latih yang cukup besar. Namun, kNN juga memiliki kelemahan, yaitu biaya komputasinya cukup tinggi.

V. SIMPULAN DAN SARAN

Dalam penelitian ini, dihasilkan sebuah metode pengklasifikasian kesegaran ikan berdasarkan citra digital ikan dengan membandingkan algoritma kNN, C4.5, dan Naive Bayes. Berdasarkan hasil pengujian yang telah dilakukan dengan membandingkan ketiga algoritma tersebut, didapatkan bahwa kNN memiliki akurasi yang paling tinggi diantara algoritma lainnya. Dengan demikian algoritma kNN dinilai cocok untuk digunakan dalam klasifikasi kesegaran ikan berdasarkan citra digital ikan. Hal ini dikarenakan kNN tangguh terhadap data latih yang banyak memiliki *noise* dan efektif untuk data latih yang cukup besar.

DAFTAR PUSTAKA

- [1] Gianie, "No Title," 2015. [Online]. Available: <http://print.kompas.com/baca/2015/05/26/Konsumsi-Ikan-Negeri-Bahari-Ini-Masih-Rendah>.
- [2] "DeZyre," 2017. [Online]. Available: <https://www.dezyre.com/article/top-10-machine-learning-algorithms/202>. [Accessed: 25-Oct-2017].
- [3] F. Muhammad, "No Title." [Online]. Available: <https://ikansegar.wordpress.com/2008/08/14/kualitas-ikan-danparameternya/>.
- [4] K. Jaya, I, Ramadhan, "Aplikasi Metode Akustik untuk Uji Kesegaran Ikan," *Bul. Teknol. Has. Perikanan*, IX, 2006.
- [5] F. . Hermawati, *Pengolahan Citra Digital*. CV. Andi Offset, 2013.
- [6] J. R. Quinlan, "C4.5 programs for machine learning," 1993.
- [7] S. Ruggieri, "Efficient C4.5," *IEEE Trans. Knowl. Data Eng.*, vol. 14, no. 2, 2002.
- [8] A. McCallum, A and Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," 1998.



UMN