

Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Tingkat Kematangan Buah Tomat Berdasarkan Fitur Warna

Hasan Mubarak¹, Sylviana Murni², Mayanda Mega Santoni³
Program Studi Informatika / Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jakarta
Jl. RS. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia
Email: hasanm@upnvj.ac.id¹, sylvianam@upnvj.ac.id², megasantoni@upnvj.ac.id³

Abstrak. Penelitian ini mengusulkan pengklasifikasi tingkat kematangan buah tomat berdasarkan fitur warna menggunakan klasifikasi K-Nearest Neighbor yang bertujuan untuk memperoleh hasil akurasi yang optimal melalui pengujian terhadap nilai k. Pada penelitian ini metode yang digunakan adalah ekstraksi fitur warna RGA*. Praproses citra menggunakan metode cropping menjadi 100×100 piksel kemudian mengkonversinya menjadi citra RGB dan Lab untuk mendapat mean RGA*. Hasil dari pengujian k dengan menggunakan enam kali percobaan dengan nilai k=1, k=3, k=5, k=7, k=9, dan k=11 dapat disimpulkan bahwa akurasi nilai K-Nearest Neighbor tertinggi sebesar 88% dengan jarak *Euclidean* k=7. Berdasarkan tingkat akurasi, nilai k=7 merupakan nilai k optimal dalam pengklasifikasian tingkat kematangan buah tomat berdasarkan fitur warna.

Kata Kunci: klasifikasi, RGB, ekstraksi fitur, KNN.

1 Pendahuluan

Tomat ialah buah yang mempunyai tingkatan kematangan tertentu. Perubahan tingkat kematangan tomat terjadi dalam waktu yang relatif singkat sehingga diperlukan teknik pemilihan yang baik untuk menghindari adanya kebusukan buah [1]. Teknik pemilihan tomat biasanya dilakukan dengan melihat perbedaan warna buah. Menurut (Harllee) tingkat kematangan tomat berdasarkan warna buah dibagi menjadi enam tingkat, yaitu: green, breakers, turning, pink, light red, serta red. Maka, warna mempunyai peranan penting dalam pengklasifikasian buah tomat [2].

Dalam pengklasifikasian kematangan buah tomat pada umumnya para petani dan masyarakat Indonesia masih menggunakan metode manual. Proses klasifikasi kematangan buah dengan menggunakan teknik manual mempunyai kelemahan yaitu memerlukan proses pengklasifikasian yang lama, tingkat akurasi yang dihasilkan rendah, serta hasil yang tidak konsisten [3]. Hal tersebut terjadi karena penentuan yang dilakukan pekerja masih bersifat subjektif. Berdasarkan masalah tersebut, maka pada penelitian ini berkaitan dengan klasifikasi kematangan buah tomat secara otomatis melalui pemanfaatan pengolahan citra digital menggunakan ekstraksi fitur warna. Klasifikasi tingkat kematangan tomat secara otomatis memberikan kemudahan dan kecepatan dalam melakukan pengujian tingkat kematangan tomat, karena penentuannya dilakukan secara obyektif [3].

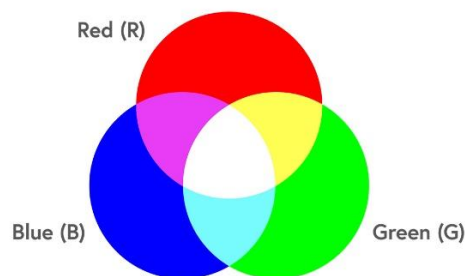
Penelitian terkait pengklasifikasian tingkat kematangan buah tomat menggunakan ekstraksi fitur warna telah dilakukan sebelumnya. Penelitian pertama yaitu klasifikasi tingkat kematangan buah tomat berdasarkan fitur warna menggunakan multi-SVM dan Penelitian kedua klasifikasi tingkat kematangan buah tomat berdasarkan fitur warna menggunakan KNN. Terdapat perbedaan penggunaan nilai rata-rata warna dan metode klasifikasi pada penelitian tersebut. Penelitian yang pertama menggunakan nilai rata-rata RGA* dan menggunakan metode klasifikasi multi-SVM hasil akurasi sebesar 77,84% [1]. Sementara itu pada penelitian kedua menggunakan nilai rata-rata RGB dan metode klasifikasi KNN hasil akurasi sebesar 98% [3].

Adapun tujuan penelitian ini adalah memperoleh hasil akurasi yang optimal dalam mengklasifikasikan tingkat kematangan buah tomat menggunakan ekstraksi fitur warna RGA* dan menggunakan metode klasifikasi KNN dengan melakukan pengujian terhadap nilai k.

2 Landasan Teori

2.1 Warna RGB

RGB merupakan ruang warna yang tersusun dari tiga channel warna yakni channel merah (Red), channel hijau (Green), dan channel biru (Blue). Setiap channel mempunyai intensitas warna yang berbeda, nilai intensitas warna minimum adalah nol (0) dan nilai intensitas warna maksimum adalah 255 (8 bit). Setiap piksel citra RGB memiliki variasi warna sejumlah 16.777.216 ($256 \times 256 \times 256$) [4]. Representasi warna RGA dapat dilihat pada Gambar 1 berikut.

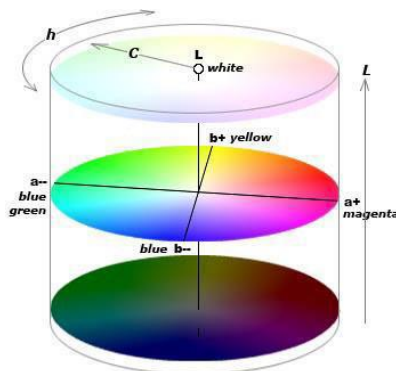


Gambar. 1. Warna RGB

2.2 Warna Lab

Warna Lab dikenal sebagai $CIE L^*a^*b^*$ (CIELAB) yang dapat dilihat pada Gambar 2. CIELAB adalah model tiga dimensi, yang hanya dapat dideskripsikan dalam ruang tiga dimensi. Diagram chromaticity dibentuk berdasarkan irisan antara komponen a^* dan komponen b^* [5]. Setiap dimensi yang terbentuk dari warna CIELAB dapat memberikan makna [6]:

- Besaran $CIE L^*$ menggambarkan kecerahan suatu warna, L^* bernilai 0 untuk warna hitam, dan 100 untuk warna putih.
- Dimensi $CIE a^*$ menggambarkan jenis warna hijau dan merah, nilai negatif a^* berarti warna hijau, dan $CIE a^*$ bertanda positif berarti merah.
- Dimensi $CIE b^*$ mendeskripsikan warna biru dan kuning, nilai negatif b^* berarti warna biru, dan $CIE b^*$ bernilai positif berarti warna kuning.



Gambar. 2. Warna LAB

2.3 K-Nearest Neighbor (KNN)

Algoritma K-Nearest Neighbor (KNN) merupakan metode pengklasifikasian kumpulan data pembelajaran berdasarkan jarak yang paling dekat dengan object [7]. Algoritma KNN mengelompokkan data baru sesuai dengan nilai k sebagai jarak tetangga terdekat antara data latih dengan data uji. Jauh atau dekatnya tetangga dapat dihitung menggunakan jarak euclidean dengan rumus [8]:

$$d(x, y) = \sqrt{\sum_{i=1}^p (x_i - y_i)^2} \quad (1)$$

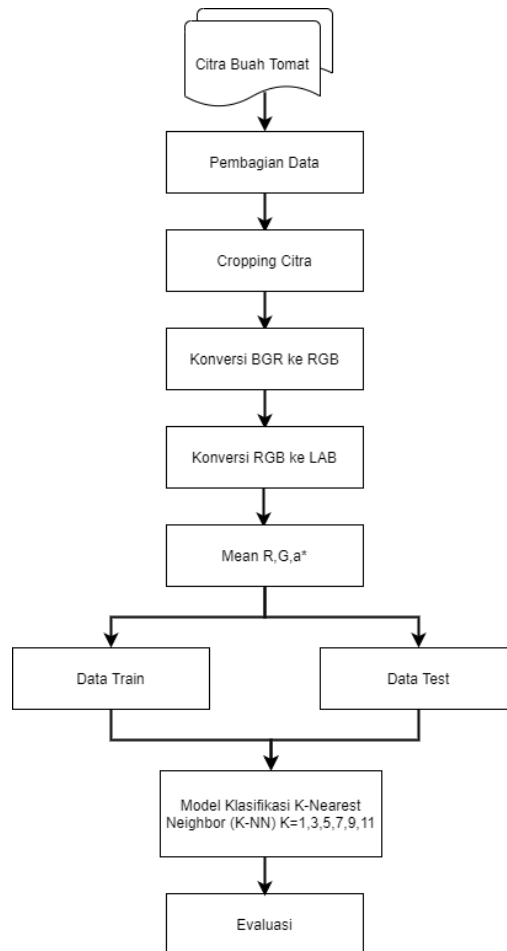
Keterangan :

$d(x,y)$: Jarak data latih dan uji,
 x_i : Data latih,
 y_i : Data uji,
 i : Variabel data,
 p : Dimensi data.

3 Metodologi Penelitian

3.1 Tahapan Penelitian

Tahapan penggunaan algoritma KNN untuk mengklasifikasikan kematangan buah tomat berdasarkan ciri warna meliputi lima tahapan yaitu pengumpulan data citra buah tomat, preprocessing, ekstraksi ciri, klasifikasi dan evaluasi hasil penelitian. Proses tahapan di gambarkan pada Gambar 3.



Gambar. 3. Tahapan Penelitian

3.1.1 Citra Tomat

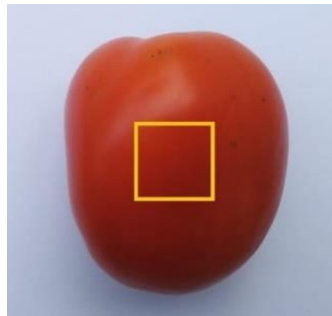
Dalam penelitian ini citra yang digunakan adalah citra buah tomat biasa (*Lycopersicum commune*) dengan total citra yang digunakan sebanyak 90 citra yang dibagi menjadi 3 jenis tomat menurut tingkat kematangannya yaitu belum matang, setengah matang dan matang. Masing-masing tentunya memiliki 30 citra buah tomat dalam format JPG dengan dimensi 413 x 427 piksel. Berikut sampel buah tomat pada Gambar 4.



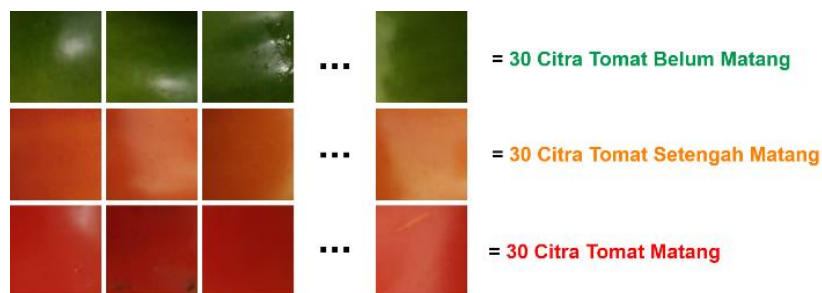
Gambar. 4. Sample Citra Buah Tomat

3.1.2 Cropping Citra

Tujuan dari cropping citra ini adalah untuk mendapatkan warna yang lebih spesifik dari citra tomat. Cropping dilakukan dengan memotong citra tomat dari bagian pusat citra menggunakan square cropping dengan ukuran citra menjadi 100x100 pixel. Hasil cropping sample citra dapat dilihat pada Gambar 5 dan Gambar 6.



Gambar. 5. Proses Pemotongan Citra Tomat



Gambar. 6. Hasil Cropping Citra Tomat

3.1.3 Konversi BGR ke RGB

Citra yang telah disegmentasi selanjutnya citra akan dikonversi menjadi citra RGB dengan menggunakan fungsi `cv.cvtColor` yang terdapat pada library `opencv`. Tujuan mengkonversi citra ini untuk kita mendapat nilai Red dan Green.



Gambar. 7. Hasil konversi citra BGR ke RGB

3.1.4 Konversi RGB ke LAB

Citra yang sebelumnya telah di konversi ke RGB akan kita gunakan untuk mendapat nilai a^* dengan mengkonversinya menjadi LAB menggunakan fungsi `cv.cvtColor` yang terdapat pada library `opencv`.



Gambar. 8. Hasil konversi citra RGB ke LAB

3.1.5 Mean R, G, a^*

Ekstraksi ciri yang dipilih adalah ekstraksi ciri warna karena indikator dalam menentukan tingkat kematangan buah tomat adalah warna dan komponen warna yang digunakan ialah warna RGa^* . Dari citra RGB didapatkan komponen warna R dan G, dan dari citra LAB didapatkan komponen warna a^* . Untuk mendapat nilai rata-rata pada masing-masing komponen dapat menggunakan fungsi `mean()`. Tabel 1 menunjukkan perhitungan komponen warna pada citra tomat.

Tabel 1. Hasil Ekstraksi ciri nilai rata-rata RGa^*

Citra Tomat	Komponen Warna		
	R	G	a^*
tbm1	67.5532	86.1949	110.5684
tbm2	40.4412	59.8579	111.3572
tbm3	58.9578	77.6096	111.693
tbm4	60.0149	79.0283	110.7363
tbm5	67.3163	85.2301	111.2112
tsm1	160.0636	61.5226	166.6992
tsm2	191.9492	89.4793	166.2427
tsm3	176.6844	58.3126	174.4656
tsm4	186.114	111.5296	152.1048
tsm5	193.1214	73.4672	173.4957
tm1	177.9773	44.6183	179.7938
tm2	139.8302	24.4295	174.3993
tm3	150.6352	21.1405	178.6753
tm4	149.558	21.5494	178.2516
tm5	163.1798	45.9157	173.9388

3.1.6 Pembagian Data

Pembagian data dibagi menjadi dua yakni data latih dan data uji. Dalam penelitian ini pembagian data sebesar 70% data latih dan 30% data uji. Berdasarkan Tabel 1, terdapat 90 data yang terbagi menjadi 63 data latih dan 27 data uji menggunakan metode *stratified random sampling*. Data latih digunakan untuk pemilihan fitur dan kemudian digunakan untuk membangun model klasifikasi. Sedangkan data uji digunakan untuk menguji model sistem klasifikasi.

3.1.7 Klasifikasi dengan K-Nearest Neighbor

Nilai fitur warna RGa* pada Tabel 1 akan digunakan untuk proses klasifikasi dengan menggunakan K-Nearest Neighbor. Pada penelitian ini akan membandingkan nilai akurasi berdasarkan nilai k yang digunakan. Pengujian akan dilakukan dengan 6 kali percobaan, yaitu dengan k=1, k=3, k=5, k=7, k=9, k=11.

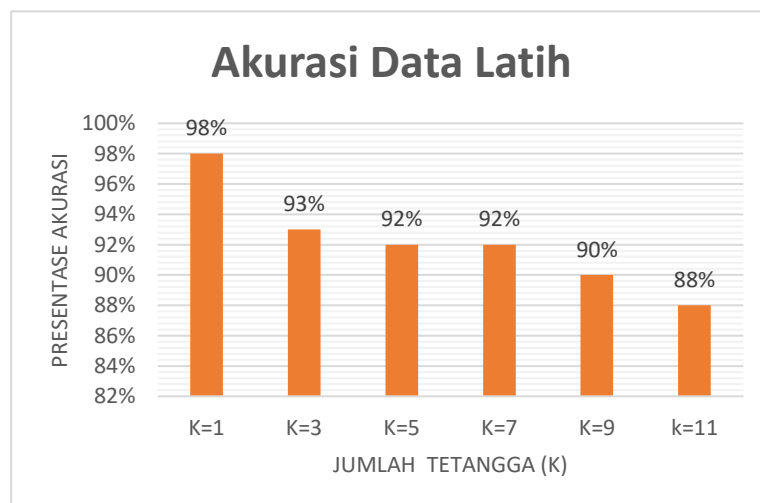
3.1.8 Evaluasi Hasil

Tahap akhir dari penelitian ini adalah menghitung akurasi hasil klasifikasi buah tomat. Hasil penelitian ini diukur dengan pengujian dan keakuratan data citra yang diamati. Tingkat akurasi dapat dinyatakan dengan rumus berikut:

$$\text{Tingkat akurasi} = \frac{\text{Jumlah tomat terklasifikasi dengan benar}}{\text{jumlah total data uji tomat}} \times 100 \% \quad (2)$$

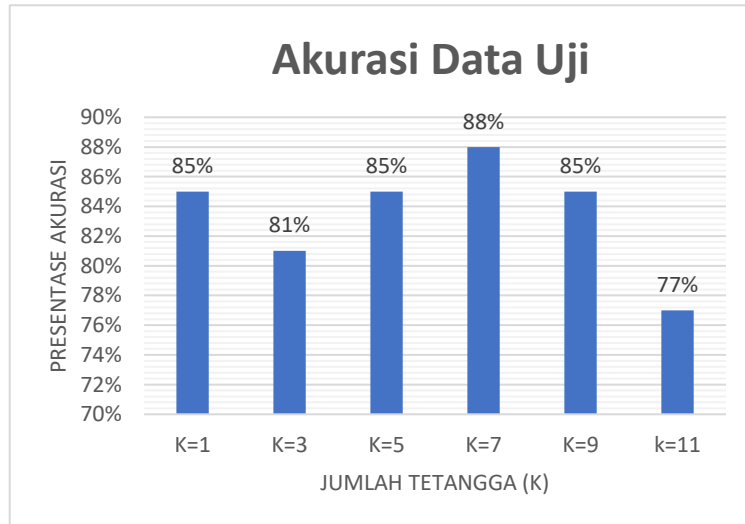
4 Hasil dan Pembahasan

Pengujian pada penelitian ini dilakukan dengan 6 kali percobaan, yaitu dengan k=1, k=3, k=5, k=7, k=9, dan k=11. Untuk memperoleh nilai k yang optimal diperlukannya perhitungan akurasi dengan menggunakan data latih. Berikut ini hasil pengujian akurasi dengan menggunakan data latih.



Gambar. 9. Hasil Akurasi Data Latih

Hasil pegujian nilai K-Nearest Neighbor dengan menggunakan data latih menunjukkan nilai akurasi tertinggi yaitu 98% terdapat pada nilai k=1, sehingga sistem dapat dikatakan layak untuk digunakan. Berikut grafik hasil akurasi menggunakan data uji.



Gambar. 10. Hasil Akurasi Data Uji

Berdasarkan Gambar 10 nilai k=7 menghasilkan akurasi tertinggi yaitu 88%. Dengan menggunakan nilai k = 7 dan jumlah data uji citra sebanyak 27 data, hasil pengujian didapatkan output yang sesuai dengan target adalah 24, dan output yang tidak sesuai target adalah 3, maka tingkat akurasinya sebesar 88%. Berikut tabel hasil pengujian dengan menggunakan k = 7.

Tabel 2. Hasil Pengujian dengan k=7

Data Pengujian			
No citra uji	Belum Matang	Setengah Matang	Matang
22	√	√	√
23	√	√	√
24	√	√	√
25	√	√	√
26	√	√	√
27	√	√	-
28	√	√	√
29	√	√	-
30	√	√	-
% Berhasil	100%	100%	68%
% Gagal	0.00%	0.00%	32%

Dari hasil perhitungan confusion matrix yang dilakukan pada proses klasifikasi dengan metode K-Nearest Neighbor, maka dihasilkan ringkasan nilai sebagai berikut:

Tabel 3. Tabel Confussion Matrix Pengujian dengan k=7

Aktual	Prediksi		
	Matang	Setengah Matang	Belum Matang
Matang	6	3	0
Setengah Matang	0	9	0
Belum Matang	0	0	9

Berdasarkan hasil perhitungan confusion matrix pada Tabel 3 terdeteksi *True-Positive* sebesar 24 data, *False-Negative* sebesar 3 data, dan *False-Positive* sebesar 3 data, menghasilkan nilai *accuracy* 92%, *precision* 88%, dan *recall* 92%.

5 Penutup

5.1 Kesimpulan

Dataset dalam penelitian ini menggunakan 90 citra buah tomat yang terbagi menjadi 63 data latih dan 27 data uji. Perbaikan citra pada tahap preprocessing dilakukan dengan metode cropping menjadi ukuran 100 x 100 pixel. Dalam ekstraksi ciri digunakan tiga ciri warna yang mempengaruhi kematangan buah tomat, yaitu RGa*. Penelitian ini menggunakan metode klasifikasi K-Nearest Neighbor dengan menguji nilai k untuk mendapatkan hasil akurasi terbaik. Nilai k optimum pada pengujian akurasi nilai K-Nearest Neighbor adalah k=7 dengan memiliki nilai akurasi tertinggi sebesar 88%. Dari 27 data citra uji diperoleh hasil akurasi terbaik sebesar 88% dengan output yang sesuai target sejumlah 24 dan yang tidak sesuai sejumlah 3 data. Penyebab kesalahan dalam penelitian ini berasal dari sampel buah tomat setengah matang dan matang. Data citra latih pada kedua jenis tomat tersebut memiliki nilai rata-rata RGa* yang hampir sama.

5.2 Saran

Berdasarkan hasil percobaan dan simpulan pada penelitian ini, maka saran yang dapat di berikan untuk penelitian selanjutnya adalah dilakukannya penambahan data citra serta menggunakan ekstraksi ciri yang lain dan metode lainnya dalam mengklasifikasi.

Referensi

- [1] Riska, S. Y., and Subeki, P. "Klasifikasi Level Kematangan Buah Tomat Berdasarkan Fitur Warna Menggunakan Multi-SVM". *Jurnal Ilmiah Informatika* 1.1 (2016).

- [2] Harllee Packing Inc. “Product: Premium product are a Harllee tradition”. Maret from <http://www.harleepacking.com/products/>.
- [3] Aprilisa, S., and Sukemi. “Klasifikasi Tingkat Kematangan Buah Tomat Berdasarkan Fitur Warna Menggunakan *K-Nearest Neighbor*”. *Jurnal Computer Science and ICT* 5.1 (2019).
- [4] Prabowo Heru. “Deteksi Kondisi Kematangan Buah Jeruk Berdasarkan Kemiripan Warna Pada Ruang Warna RGB Berbasis Android”. *Jurnal Elektronik Sistem Informasi dan Komputer* 3.2 (2017).
- [5] Tupamahu, F., & Enggar Sukmana, S. (n.d.). “Ekstraksi Connected Component dan Transformasi Ruang Warna CIELAB Untuk Segmentasi Citra Penyakit Pada Daun Tanaman Jagung”. *Seminar Nasional Ke, 9* (2014):91-96.
- [6] Sinaga, A. S. “SEGMENTASI RUANG WARNA $L^*a^*b^*$ ”. *Jurnal Mantik Penusa* 3.1 (2019).
- [7] Andono, P. N., Sutojo, T., & Mulijono. (2017). *Pengelolaan Citra Digital*. Yogyakarta: Penerbit ANDI.
- [8] Lestari, M. “Penerapan Algoritma Klasifikasi Nearest Neighbor (K-NN) untuk Mendeteksi Penyakit Jantung”. *Faktor Exacta* 7.2 (2014).
- [9] Sanjaya, Suwanto. “Penerapan Learning Vector Quantization Pada Pengelompokan Tingkat Kematangan Buah Tomat Berdasarkan Warna Buah”. *Jurnal CoreIT* 5.2 (2019).
- [10] Yuhandri, Y. “Perbandingan Metode Cropping Pada Sebuah Citra Untuk Pengambilan Motif Tertentu Pada Kain Songket Sumatera Barat”. *Komtekinfo* 6.1 (2019).