

# Penerapan Machine Learning Untuk Klasifikasi Tingkat Kematangan Buah Anggur (*Vitis*) Dengan Metode K-Nearest Neighbor

Sri Melisa\*, Achmad Fauzi, Yusfrizal

Program Studi Sistem Informasi, STMIK Kaputama, Binjai, Indonesia

Email: <sup>1,\*</sup>srnelisa00@gmail.com, <sup>2</sup>fauzyrivai88@gmail.com, <sup>3</sup>yusfrizal80@gmail.com

Email Penulis Korespondensi: srnelisa00@gmail.com

**Abstrak**-Anggur merupakan tanaman buah berupa perdu merambat yang termasuk ke dalam keluarga Vitaceae yang dapat dimakan langsung atau diolah menjadi minuman dan makanan. Buah anggur memiliki ciri warna untuk menentukan tingkat kematangan buah secara benar. Penentuan kematangan buah anggur biasanya dilakukan langsung dengan melihat warna kematangan dari buah anggur yang dapat dilihat dari warna merah jika sudah matang dan hijau jika belum matang. Dan hal ini bisa saja terjadi kekeliruan dalam mengelompokkan dan menentukan tingkat kematangan buah. Klasifikasi merupakan cara pengelompokkan benda berdasarkan ciri-ciri yang dimiliki oleh objek klasifikasi. Klasifikasi yang dilakukan oleh komputer menggunakan metode yang digunakan untuk melakukan klasifikasi. K-nearest neighbor adalah salah satu metode yang dapat digunakan dalam klasifikasi dan merupakan algoritma machine learning dengan pendekatan supervised learning yang bekerja dengan mengelompokkan data baru menggunakan kemiripan antara data baru dengan sejumlah data (k) pada lokasi yang terdekat yang telah tersedia. Algoritma KNN digunakan untuk klasifikasi dan regresi. Dengan menggunakan kategori mayoritas, maka dapat hasil klasifikasi dari ketiga data terbaik tersebut diberi label target sesuai dengan dataset awal. Data terbaik adalah data ke 2, 3, 10 dari data tersebut terdapat 1 kategori belum matang dan 2 matang, sehingga mayoritas dari hasil klasifikasi tersebut yaitu buah anggur matang. Jadi hasil klasifikasi untuk data citra uji 11 yaitu buah anggur masuk dalam kategori klasifikasi matang.

**Kata Kunci:** Anggur; Klasifikasi; KNN; Machine Learning; Warna.

**Abstract**-Grapes are fruit plants in the form of vines belonging to the Vitaceae family that can be eaten directly or processed into drinks and food. Grapes have color characteristics to determine the level of ripeness of fruit correctly. Determination of ripeness of grapes is usually done directly by looking at the color of ripeness of the fruit grapes that can be seen from the color red if it is ripe and green if it is not ripe. And this could be a mistake in classifying and determining the level of fruit maturity. Classification is a way of grouping objects based on the characteristics possessed by the object of classification. Classification is carried out by computers using the methods used to classify. K-nearest neighbor is a method that can be used in classification and is a machine learning algorithm with a supervised learning approach that works by classifying new data using the similarity between new data and a number of data (k) at the closest available location. KNN algorithm is used for classification and regression. By using the majority category, the classification results from the three best data can be labeled as targets according to the initial dataset. The best data is data to 2, 3, 10 of these data, there are 1 category of immature and 2 ripe, so that the majority of the classification results are ripe grapes. So the classification results for test image data 11 are grapes included in the ripe classification category.

**Keywords:** Classification; Color; KNN; Machine Learning; Wine.

## 1. PENDAHULUAN

Anggur merupakan tanaman buah berupa perdu merambat yang termasuk ke dalam keluarga *Vitaceae*. Buah ini biasanya digunakan untuk membuat jus anggur, jelly, minuman anggur, minyak biji anggur dan kismis, atau dimakan langsung. Buah ini juga dikenal karena mengandung banyak senyawa *polifenol* dan *resveratol* yang berperan aktif dalam berbagai metabolisme tubuh, serta mampu mencegah terbentuknya sel kanker dan berbagai penyakit lainnya. Aktivitas ini juga terkait dengan adanya senyawa metabolit sekunder di dalam buah anggur yang berperan sebagai senyawa antioksidan yang mampu menangkal radikal bebas.

Klasifikasi merupakan cara pengelompokkan benda berdasarkan ciri-ciri yang dimiliki oleh objek klasifikasi. Dalam prosesnya, klasifikasi dapat dilakukan dengan banyak cara baik secara manual ataupun dengan bantuan teknologi. Klasifikasi yang dilakukan secara manual adalah klasifikasi yang dilakukan oleh manusia tanpa adanya bantuan dari algoritma cerdas komputer. Sementara klasifikasi yang dilakukan oleh komputer yaitu menggunakan metode-metode yang digunakan untuk melakukan klasifikasi.

Buah anggur tentu memiliki ciri warna untuk menentukan tingkat kematangan buah anggur secara benar. Dari penelitian yang dilakukan untuk menentukan kematangan buah anggur biasanya dilakukan langsung dengan melihat warna kematangan dari buah anggur yang dapat dilihat dari warna merah jika sudah matang dan hijau jika belum matang. Hal ini tentu bisa saja terjadi kekeliruan dalam mengelompokkan dan menentukan tingkat kematangan buah anggur.

Untuk itu penelitian ini ingin melakukan klasifikasi tingkat kematangan buah anggur dengan menggunakan komputer dan penerapan sebuah metode yang dapat digunakan untuk klasifikasi tingkat kematangan buah anggur. Metode yang akan digunakan dalam penelitian ini yaitu menggunakan metode KNN (K-Nearest Neighbor). *K-nearest Neighbor* adalah salah satu algoritma machine learning dengan pendekatan supervised learning yang bekerja dengan mengelompokkan data baru menggunakan kemiripan antara data baru dengan sejumlah data (k) pada lokasi yang terdekat yang telah tersedia. Algoritma KNN digunakan untuk klasifikasi dan regresi. KNN juga merupakan metode klasifikasi yang sangat sederhana dalam mengklasifikasikan sebuah gambar berdasarkan jarak terdekat dengan tetangganya.

Penerapan metode KNN telah banyak dilakukan oleh para peneliti diantaranya dilakukan oleh tentang implementasi *K-Nearest Neighbor* untuk klasifikasi bunga dengan ekstraksi fitur warna RGB. Hasil dari penerapan

metode ini metode klasifikasi KNN dengan metode ekstraksi fitur warna RGB kurang cocok untuk mengklasifikasi label objek dengan kemiripan warna walaupun bentuknya berbeda namun masih bisa dipakai dalam mengklasifikasikan objek dengan perbedaan warna yang mencolok di salah satu sisi walaupun bentuk objeknya memiliki struktur bentuk yang mirip. Untuk meingkatkan akurasi rata-rata klasifikasi label pada bunga yang memiliki kesamaan warna perlu dilakukan uji coba dengan metode klasifikasi KNN dengan tambahan ekstraksi fitur dan parameter lainnya atau metode lain yang lebih handal [1].

Peneliti selanjutnya oleh tenang klasifikasi kematangan buah mangga berdasarkan citra HSV dengan KNN. Kesimpulan dari penelitian ini metode *K-NN* dapat diterapkan dengan baik dalam sistem klasifikasi kematangan mangga berdasarkan warna HSV. Akurasi yang didapatkan dari pengujian data testing memiliki rata-rata akurasi 55 % dengan jarak antara  $k=1-10$ . Menggunakan data *training* sebanyak 129 mangga dan data *testing* sebanyak 40 mangga dengan masing-masing kelas yakni kelas mentah, kelas cukup, kelas matang, dan kelas sangat matang [2].

## 2. METODOLOGI PENELITIAN

### 2.1 Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dengan kata lain, klasifikasi adalah proses pengumpulan data bersama yang didasarkan atas sekumpulan kesamaan yang awalnya telah ditentukan oleh seorang analis sebelum analisa dimulai. Teknik ini memeriksa data yang telah diklasifikasikan dan dikumpulkan dalam group bersama-sama sesuai dengan keanggotaannya. Aturan keanggotaan bisa mempunyai komponen waktu, komponen geografis dan komponen kualitatif [3]. Contohnya:

- Untuk mencari retensi pelanggan, dibuat group pelanggan yang diklasifikasikan dan diperiksa untuk mengetahui pembelian produk sepatu atau sandal dengan harga berapapun.
- Untuk mengetahui pengaruh diskon atau potongan harga dari produk tertentu menarik perhatian pelanggan setia untuk membelinya.

Proses klasifikasi dapat dibagi menjadi tiga tahap :

- Learning. Algoritma yang mencari sejumlah *record* dari *training set* dan menciptakan sebuah deskripsi tentang model klasifikasi. Model ini dibuat kecil dan persis pada saat yang sama.
- Model yang sudah dibuat harus diuji dengan serangkaian uji coba dalam sebuah basis data. Sebagai *training set*, *record* uji coba harus merupakan kumpulan yang lebih dahulu diklasifikasikan. *Record* untuk uji coba harus berbeda dengan dengan *training set*. Model *classification* akan selalu bekerja dengan sempurna dalam *training set*.
- Classification*, model iterasi yang digunakan untuk membuat klasifikasi pada sisa *record-record* dalam basis data yang ada.

### 2.2 K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* (KNN) 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, di mana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Sebuah titik pada ruangan ini ditandai dengan kelas  $c$ , jika kelas  $c$  adalah klasifikasi yang paling banyak ditemui pada  $k$  buah tetangga terdekat titik tersebut [4].

*K-Nearest Neighbor* metode yang bersifat *supervised*, di mana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas kategori pada KNN. KNN adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumnya. Termasuk dalam *supervised learning*, di mana hasil *query instance* yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam KNN.

*K-Nearest Neighbor* (KNN) termasuk kelompok *instance based learning*. algoritma ini juga adalah salah satu teknik *lazy learning*. KNN dilakukan dengan mencari kelompok  $k$  objek dalam data *training* yang paling dekat (mirip) dengan objek pada data baru atau data *testing*. Algoritma *K-Nearest Neighbor* adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. *Nearest Neighbor* adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dan kasus lama yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada.

Untuk mendefinisikan jarak antara dua titik yaitu titik pada data training ( $x$ ) dan titik pada data testing ( $y$ ) maka dipakai rumus *Euclidean*, seperti yang ditunjukkan pada persamaan:

$$D(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (1)$$

Dengan  $D$  adalah jarak antara titik pada data training  $x$  dan titik data *testing*  $y$  yang akan diklasifikasi, di mana  $x=x_1, x_2, \dots, x_i$  &  $y=y_1, y_2, \dots, y_i$  dan  $i$  merepresentasikan nilai atribut serta  $n$  adalah dimensi atribut.

### 2.3 Anggur

Anggur merupakan tanaman buah berupa perdu merambat yang termasuk ke dalam keluarga *Vitaceae*. Buah ini biasanya digunakan sebagai menciptakan jus anggur, jelly, minuman anggur, minyak biji anggur, dan kismis, atau dimakan

langsung. Buah ini juga dikenal karena mengandung banyak senyawa polifenol dan resveratol yang berperan aktif dalam berbagai metabolisme tubuh, serta dapat mencegah terbentuknya sel kanker dan berbagai penyakit pautannya. Kegiatan ini juga terkait dengan hal hadir senyawa metabolit sekunder di dalam buah anggur yang berperan sbg senyawa antioksidan yang dapat menangkal radikal lepas sama sekali.









### 3. HASIL DAN PEMBAHASAN




Adapun langkah-langkah untuk menghitung metode Algoritma *K-Nearest Neighbor* yaitu:

- Menentukan Parameter *K* (Jumlah tetangga paling dekat).
- Menghitung kuadrat jarak *Euclid* (*query instance*)
- masing-masing objek terhadap data *sampel* yang diberikan.
- Kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak *Euclid* terkecil.
- Mengumpulkan kategori *Y* (Klasifikasi *Nearest Neighbor*)
- Dengan memakai kategori *Nearest Neighbor* yang paling mayoritas maka dapat diprediksi nilai *query instance* yang telah dihitung.

Berikut merupakan contoh perhitungan metode *K-Nearest Neighbor* dengan data hasil ekstrasi ciri citra yang diperoleh dari seperti pada tabel dibawah ini.

**Tabel 1.** Dataset Citra Buah Anggur

No.	Citra	Keterangan
1		Belum Matang
2		Belum Matang
3		Matang
4		Matang
5		Belum Matang
6		Belum Matang
7		Belum Matang
8		Belum Matang

No.	Citra	Keterangan
9		Belum Matang
10		Matang
11		?

Dari data citra di atas sebelum dilakukan analisa dengan menggunakan metode *K-Nearest Neighbor*, maka perlu dilakukan ekstraksi ciri citra dengan menggunakan ekstraksi ciri citra RGB, HSV dan area. Untuk mencari ciri citra tersebut dilakukan dengan menggunakan *software* Matlab. Sebagai contoh data citra diatas memiliki ciri-ciri sebagai berikut :

**Tabel 2.** Data Ekstraksi Ciri Citra Buah Anggur

No	Red	Green	Blue	Hue	Saturation	Value	Area	Target
Citra 1	93,76155	111,2652	46,47034	0,209771	0,643122	0,436598	395138	Belum Matang
Citra 2	86,92599	94,93308	36,27754	0,187663	0,694736	0,374155	665020	Belum Matang
Citra 3	164,7719	68,33548	61,72901	0,403677	0,662464	0,646175	493023	Matang
Citra 4	141,7829	74,15678	86,39493	0,925142	0,52645	0,556023	374982	Matang
Citra 5	94,6524	108,3329	37,76695	0,186026	0,652106	0,430293	336008	Belum Matang
Citra 6	94,6524	108,3329	37,76695	0,186026	0,652106	0,430293	336008	Belum Matang
Citra 7	86,17046	104,052	28,21793	0,204316	0,831919	0,408727	348356	Belum Matang
Citra 8	122,5048	128,905	68,01217	0,176164	0,528121	0,507932	308022	Belum Matang
Citra 9	93,76155	111,2652	46,47034	0,209771	0,643122	0,436598	395138	Belum Matang
Citra 10	130,1897	83,12062	95,73888	0,923544	0,367792	0,511504	509928	Matang
Citra 11	167,1202	69,61055	67,30931	0,520537	0,642477	0,655403	391409	?

Dataset citra buah anggur di mana terdapat 10 *record* untuk data *training* yaitu citra 1-10 dan 1 data *testing* yaitu data ke-11. Terdapat 6 atribut yaitu ekstraksi ciri citra *Red, Green, Blue, Hue, Saturation, Value* dan Area yang tipe datanya numerik (bisa dijumlahkan) dan terdapat 1 label yaitu klasifikasi yang mempunyai jenis data nominal yaitu belum matang atau matang.

Setelah ditentukan data training maka akan dilakukan proses klasifikasi pada data *testing* citra 11 apakah buah anggur tersebut masuk kedalam kelas buah anggur belum matang atau matang, dengan proses sebagai berikut.

- Menentukan parameter K, di sini dalam penentuan paramter  $k=300000$ .
- Menghitung kuadrat jarak euclid (euclidean distance) masing-masing objek terhadap data sampel yang diberikan.

Perhitungan untuk data ke-1 terhadap data ke-6 (data testing) Dari hasil perhitungan yang telah dilakukan di peroleh nilai distance sebagai berikut:

**Tabel 3.** Object dan Distance

No	Object	Distance
1	C1, C11	3.730,01
2	C2, C11	273.611,01
3	C3, C11	101.614,00
4	C4, C11	16.427,03
5	C5, C11	55.401,07
6	C6, C11	55.401,07
7	C7, C11	43.053,11
8	C8, C11	83.387,03
9	C9, C11	3.730,01
10	C10, C11	118.519,01

Menguruntukan objek-objek tersebut ke dalam kelompok jarak Euclidean terkecil untuk yang mempunyai mengurutkan kita hanya perlu membuat urutan dari data yang mempunyai jarak terkecil ke terbesar. Berikut hasil urutan berupa rangking.

**Tabel 4.** Jarak Eucliden Terkecil

No	Object	Distance	Ranking
1	C1, C11	3.730,01	8
2	C2, C11	273.611,01	1
3	C3, C11	101.614,00	3
4	C4, C11	16.427,03	10
5	C5, C11	55.401,07	5
6	C6, C11	55.401,07	6
7	C7, C11	43.053,11	7
8	C8, C11	83.387,03	4
9	C9, C11	3.730,01	9
10	C10, C11	118.519,01	2

Mengumpulkan kategori Y (klasifikasi nearest neighbor), Pada tahap ini hanya mengambil data sesuai dengan jumlah k yang telah ditentukan di langkah 1, Pada langkah 1, k yang kita tentukan adalah k=300000, jadi kita memilih 3 data terbaik saja. Hasilnya yaitu sebagai berikut:

**Tabel 5.** Rangking

No	Object	Distance	Ranking
1	C1, C11	3.730,01	8
2	C2, C11	273.611,01	1
3	C3, C11	101.614,00	3
4	C4, C11	16.427,03	10
5	C5, C11	55.401,07	5
6	C6, C11	55.401,07	6
7	C7, C11	43.053,11	7
8	C8, C11	83.387,03	4
9	C9, C11	3.730,01	9
10	C10, C11	118.519,01	2


Data 3 terbaik dari tabel di atas yaitu nomor 2, 3 dan 10 dan dapat dilihat seperti pada tabel dibawah ini.

**Tabel 6.** Klasifikasi Data Terbaik

No	Nama	Distance	Ranking	Target
2	C2, C11	273.611,01	1	Belum Matang
3	C3, C11	101.614,00	3	Matang
10	C10, C11	118.519,01	2	Matang

Dengan menggunakan kategori mayoritas, maka dapat hasil klasifikasi dari ketiga data terbaik tersebut diberi label target sesuai dengan dataset awal. Data terbaik adalah data ke 2, 3, 10 dari data tersebut terdapat 1 kategori belum matang dan 2 matang, sehingga mayoritas dari hasil klasifikasi tersebut yaitu buah anggur matang. Jadi hasil klasifikasi untuk data citra uji 11 yaitu buah anggur masuk dalam kategori klasifikasi matang.

**Tabel 7.** Hasil Klasifikasi

Citra	Ekstraksi Ciri Citra	Hasil Klasifikasi
	Red	167,1202
	Green	69,61055
	Blue	67,30931
	Hue	0,5205
	Saturation	0,6425
	Value	0,6554
	Area	391409
		<b>Matang</b>

#### 4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, maka kesimpulan yang dapat diambil yakni Sistem dapat memberikan informasi tingkat kematangan buah anggur berdasarkan warna dari buah anggur. Sistem dapat membantu mengurangi kekeliruan para petani anggur dalam mengelompokkan dan menentukan tingkat kematangan buah anggur. Sistem dapat mengefisiensi waktu para petani dalam menentukan tingkat kematangan buah yang biasanya dilakukan secara manual.

## REFERENCES

- [1] Farokhah, L. (2020). Implementasi K-Nearest Neighbor untuk Klasifikasi Bunga Dengan Ekstraksi Fitur Warna RGB. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(6), 1129.
- [2] Nafiah, N. (2019). Klasifikasi Kematangan Buah Mangga Berdasarkan Citra HSV dengan KNN. *Jurnal Elektronika Listrik Dan Teknologi Informasi Terapan*, 1(2), 1–4.
- [3] Relita Buaton, Muhammad Zarlis, Syahril Efendi, V. Y. (2019). *Data Mining*. Wade Group, Purwosari.
- [4] Amril Mutoi Siregar, A. P. (2016). *DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner*. CV Kekata Group, Suko Harjo.
- [5] Away, G. A. (2014). *The Shortcut of MATLAB Programming*. Informatika. Bandung.
- [6] Permadi, Y., & Murinto, . (2015). Aplikasi Pengolahan Citra Untuk Identifikasi Kematangan Mentimun Berdasarkan Tekstur Kulit Buah Menggunakan Metode Ekstraksi Ciri Statistik. *Jurnal Informatika*, 9(1), 1028–1038.
- [7] Prahudaya, T. Y., & Harjoko, A. (2017). Metode Klasifikasi Mutu Jambu Biji Menggunakan Knn Berdasarkan Fitur Warna Dan Tekstur. *Jurnal Teknosains*, 6(2), 113.
- [8] Pulung Nurtantio Andono, T.Sutojo, M. (2017). *Pengolahan Citra Digital*. CV. Andi Offset, Yogyakarta.
- [9] Pusadan, M. Y. (2015). *Pemrograman MATLAB pada Sistem Pakar Fuzzy*. CV. Budi Utama. Yogyakarta.
- [10] Sitorus, L. (2015). *Algoritma dan Pemograman*. CV. Andi Offset, Yogyakarta.
- [11] Sugiarti, Y. (2013). *Analisis dan Perancangan UML (United Modeling Language) Generated VB.6*. Graha Ilmu. Yogyakarta.