

*Skripsi*

**IDENTIFIKASI KUALITAS APEL *GOLDEN* MELALUI WARNA CITRA  
MENGUNAKAN ALGORITMA *K-MEANS CLUSTERING***

**FITRI**

**H021 19 1071**



**DEPARTEMEN FISIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR**

**2022**

**IDENTIFIKASI KUALITAS APEL *GOLDEN* MELALUI WARNA CITRA  
MENGUNAKAN ALGORITMA *K-MEANS CLUSTERING***

**SKRIPSI**

*Diajukan Sebagai Salah Satu Syarat Memperoleh Gelar  
Sarjana Sains pada Program Studi Fisika Departemen Fisika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Hasanuddin*

**FITRI**

**H021 19 1071**

**DEPARTEMEN FISIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS HASANUDDIN**

**MAKASSAR**

**2022**

**LEMBAR PENGESAHAN**

**IDENTIFIKASI KUALITAS APEL *GOLDEN* MELALUI WARNA CITRA  
MENGUNAKAN ALGORITMA *K-MEANS CLUSTERING***

**Disusun dan diajukan oleh:**

**FITRI**

**H021 19 1071**

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka  
Penyelesaian Studi Program Sarjana Program Studi Fisika  
Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin  
pada tanggal 09 November 2022  
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,



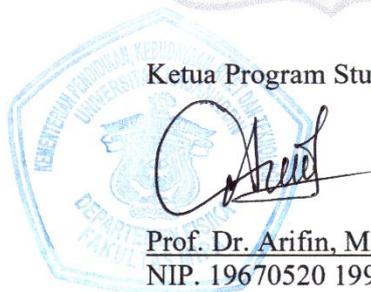
Prof. Dr. rer.nat. Wira Bahari Nurdin  
NIP. 19670923 199003 1 001

Pembimbing Pertama,



Eko Juarlín, S.Si., M.Si.  
NIP. 19811106 200812 1 002

Ketua Program Studi,



Prof. Dr. Arifin, M.T.  
NIP. 19670520 199403 1 002

## PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Fitri  
NIM : H021 19 1071  
Program Studi : Fisika  
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya yang berjudul

### **Identifikasi Kualitas Apel *Golden* Melalui Warna Citra Menggunakan Algoritma *K-Means Clustering***

Adalah karya tulisan saya sendiri dan bukan merupakan pengambil alihan tulisan orang lain dan bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau seluruh skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 19 November 2022  
Yang Menyatakan,



## ABSTRAK

Kemajuan teknologi memungkinkan identifikasi kualitas buah apel dengan bantuan komputer. Namun, identifikasi kualitas buah secara otomatis menggunakan komputer masih merupakan tugas menantang karena terdapat beragam jenis buah-buahan khususnya buah apel. Pada penelitian ini, kami membuat program untuk mengidentifikasi kualitas buah apel khususnya apel *golden* melalui aspek warna citra menggunakan metode *k-means clustering*. Terdapat 300 data citra dengan tiga atribut yang digunakan dalam penelitian ini. Variasi pemilihan *centroid* awal dipilih sebanyak tujuh variasi yaitu A, B, C, D, E, F dan G. Penelitian ini mampu membuat program *machine learning* yang dapat mengidentifikasi kualitas buah apel *golden* ke dalam tiga *cluster* berdasarkan indeks warna citra yaitu merah, hijau dan biru (RGB). Hasil yang diperoleh menunjukkan bahwa metode yang diusulkan dapat digunakan untuk mengidentifikasi citra apel *golden* dengan tingkat akurasi sebesar 95,33%, masing-masing untuk variasi A, D, E, F, G dan 66,67% untuk variasi B dan C.

***Kata Kunci:*** *Machine learning; K-Means Clustering; Apel Golden; RGB*

## **ABSTRACT**

Advancement in technology allow us to identify different quality of fruit using computer. However, automatic identification of fruit variant using a computer remain a challenging task because there are various type of fruits, especially apples. In this study, we made a program to identify quality of apples, especially golden apples. Through the color aspect using the k-means clustering method, we study 300 images data with three attributes. We varied the selection of the initial centroid as many as seven variations, namely A, B, C, D, E, F and G. The aim of this study is to make a machine learning program that can identify the quality of golden apples into three clusters based on the RGB (red, green and blue) color index. The result shows that the program can be used to identify golden apples images with 95.33% accuracy for A, D, E, F and G variations and 66.67% for B and C variations.

**Keywords:** Machine learning; K-Means Clustering; Golden Apple; RGB

## KATA PENGANTAR

الرَّحِيمِ الرَّحْمَنِ اللَّهُ بِسْمِ  
تُهُ وَبَرَكَاتِ اللَّهِ وَرَحْمَةً عَلَيْكُمْ السَّلَامُ

Alhamdulillah puji syukur penulis panjatkan ke hadirat Allah Subhanahu WaTa'ala penguasa alam semesta yang ditangan-Nya gudang segala urusan. Dengan qudrah-Nya kunci segala kebaikan dan kejahatan. Segala puji bagi Allah Subhanahu WaTa'ala yang Maha Pemurah memberikan karunia dan nikmat kepada hamba-Nya. Oleh karena itu, sepatutnya kita bersyukur atas segala pemberian dan nikmat-Nya, kesabaran atas semua ujian yang dibebankan-Nya yang dengannya sehingga penulis dapat menyusun skripsi ini.

Salam dan shalawat semoga senantiasa tercurah atas junjungan Rasulullah Muhammad Shallallahu'alaihi Wa sallam sebagai uswah hasanah yang telah memberi cahaya kesucian dan kebenaran hakiki kepada seluruh ummatnya dan semoga keselamatan dilimpahkan kepada seluruh keluarga dan sahabatnya maupun kedua generasi setelahnya serta para pengikutnya yang setia hingga akhir zaman.

Tidaklah mudah bagi penulis untuk dapat menyelesaikan skripsi ini. Penulis menyadari bahwa sejak penyusunan skripsi ini rampung, banyak hambatan, rintangan dan halangan. Namun semua ini dapat teratasi dengan baik berkat bantuan, motivasi dan doa dari berbagai pihak. Penulis juga menyadari bahwa skripsi ini masih jauh dari kata sempurna sehingga penulis mengharapkan kritik dan saran yang konstruktif dari pembaca demi kesempurnaan karya tulis ini. Penulis berharap dengan selesainya skripsi ini, bukanlah akhir dari sebuah karya, melainkan awal dari semuanya, awal dari sebuah perjuangan hidup, dan awal dari sebuah doa yang selalu menyertainya. Aamiin.

Penghargaan dan rasa terima kasih yang setinggi-tingginya penulis haturkan kepada:

1. Almarhumah ibu **Nur Hasanah, S.Si., M.Si.** selaku dosen yang banyak memberi kontribusi dan memberi semangat kepada penulis agar dapat menyelesaikan tugas akhir selama satu semester.

2. Bapak **Eko Juarlin, S.Si., M.Si.** selaku dosen pembimbing penulis yang telah meluangkan waktu untuk membagikan begitu banyak ilmu dan motivasi kepada penulis. Semoga beliau senantiasa diberikan kesehatan dan kekuatan untuk dapat terus membagikan ilmunya kepada para mahasiswa.
3. **Prof. Dr. rer.nat. Wira Bahari Nurdin** selaku dosen pembimbing penulis sekaligus kepala Lab. Fisika Teori dan Komputasi yang telah meluangkan waktunya untuk memberikan bimbingan dan saran yang sangat membangun dalam penyelesaian skripsi ini.
4. **Drs. Bansawang BJ, M.Si.** dan **Prof. Dr. Syamsir Dewang, M.Eng.Sc.** selaku dosen penguji yang telah memberi masukan, kritikan dan saran kepada penulis.
5. Bapak **Azwar Sutiono, S.Si., M.Si.** selaku dosen pembimbing akademik penulis yang telah memberi saran selama perkuliahan.
6. **Prof. Dr. Arifin, M.T,** selaku ketua Departemen Fisika dan seluruh **dosen** yang telah membagi ilmunya beserta **staf** Departemen Fisika FMIPA UNHAS yang telah membantu penulis selama menempuh perkuliahan hingga selesai.
7. Penghuni grup Lab. Teori 2019 (**I\_079, M\_048** dan **UA\_059** serta **AAPN\_510**[2019+]). Terima kasih telah mengukir cerita bersama selama di lab. Fisika Teori dan Komputasi.
8. **MNGY\_018;** Terima kasih atas ilmunya serta telah membantu dan berkontribusi disaat penulis belum menemukan judul skripsi.
9. **ANF\_305** dan **RR\_509;** Terima kasih atas ilmunya dan sedikit cerita pada saat penulis masih maba.
10. **S\_021** dan **AAPN\_510;** Terima kasih atas cerita, bantuan dan kebersamaannya selama di Lab, yah walau terkadang agak ngeselin sih. Mohon maaf sering merepotkan kalian. *At least*, hidupku di Lab sedikit berwarna (tidak abu-abu) hehe~
11. **FAZ\_317** dan **N\_014** yang telah berkontribusi dan memberi sedikit cerita pengalaman selama perkuliahan. **MHH\_010;** Terima kasih atas ilmunya dan sedikit cerita yang dapat menghidupkan kembali suasana di Lab. Teori.

12. **PMT\_066, FF\_047, BHN\_051 dan EM\_079 dkk (fisika 2020)** yang tidak dapat penulis susun satu persatu. Terima kasih atas cerita dan kebersamaannya selama di Lab.
13. Mahasiswa **fisika angkatan 2019** yang tidak dapat penulis susun satu persatu dan penghuni **Lab. Fisika Teori dan Komputasi**, serta kepada seluruh pihak yang telah berkontribusi selama ini.
14. Terkhusus kepada penghuni **grup buat web (INSAF)**. Terima kasih telah menjadi kawan bagi penulis dalam suka maupun duka, dan berbagi pengalaman hidup selama perkuliahan. *Thank you guys ^^*
15. Tak lupa dan yang paling penting, terima kasih yang sebesar-besarnya penulis haturkan kepada orangtua tercinta Ibunda Hikmah Lanna dan Ayahanda Nasruddin Saing, saudara-saudariku tercinta serta keluargaku yang telah memberikan segala do'a, cinta, perhatian, kasih sayang, dorongan baik moril maupun materil, dengan penuh keikhlasan serta do'a restunya yang selalu mengiringi penulis dalam setiap langkah selama menempuh pendidikan. Semoga Allah *Subhanahu Wa Ta'ala* senantiasa melimpahkan Rahmat dan Hidayah-Nya kepada kita semua.

Makassar, 29 November 2022

Fitri

## DAFTAR ISI

<b>HALAMAN SAMPUL</b> .....	<b>i</b>
<b>HALAMAN JUDUL</b> .....	<b>ii</b>
<b>LEMBAR PENGESAHAN</b> .....	<b>i</b>
<b>PERNYATAAN KEASLIAN</b> .....	<b>iv</b>
<b>ABSTRAK</b> .....	<b>v</b>
<b>ABSTRACT</b> .....	<b>vi</b>
<b>KATA PENGANTAR</b> .....	<b>vii</b>
<b>DAFTAR ISI</b> .....	<b>x</b>
<b>DAFTAR GAMBAR</b> .....	<b>xii</b>
<b>DAFTAR TABEL</b> .....	<b>xiii</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xiv</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
I.1    Latar Belakang .....	1
I.2    Rumusan Masalah .....	2
I.3    Tujuan Penelitian.....	2
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>3</b>
II.1    Apel.....	3
II.2    Pengolahan Citra .....	4
II.3    Ruang Warna.....	5
II.4 <i>Machine Learning</i> .....	6
II.4.1 <i>Supervised Learning</i> .....	6
II.4.2 <i>Unsupervised Learning</i> .....	6
II.5 <i>Clustering</i> .....	6
II.6 <i>K-Means Clustering</i> .....	6
<b>BAB III METODOLOGI PENELITIAN</b> .....	<b>8</b>
III.1    Studi Literatur .....	8
III.2    Pengumpulan Data .....	8
III.3    Alat.....	8
III.4    Pengolahan Citra .....	8
III.5    Metode <i>K-Means Clustering</i> .....	8
III.6    Evaluasi Performa Model .....	9
III.7    Bagan Alir Penelitian .....	10

III.8	Bagan Alir Simulasi .....	11
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>		<b>12</b>
IV.1	Pengolahan Citra .....	12
IV.2	Pengelompokkan Data.....	12
IV.3	Evaluasi Performa Model .....	15
<b>BAB V KESIMPULAN DAN SARAN .....</b>		<b>20</b>
V.1	Kesimpulan .....	20
V.2	Saran .....	20
<b>DAFTAR PUSTAKA .....</b>		<b>21</b>
<b>LAMPIRAN.....</b>		<b>24</b>

## DAFTAR GAMBAR

	Halaman
<b>Gambar 2.1</b> Contoh citra apel <i>golden</i> .....	3
<b>Gambar 2.2</b> Representasi citra digital dalam 2D.....	5
<b>Gambar 3.1</b> Bagan alir penelitian .....	10
<b>Gambar 3.2</b> Bagan alir simulasi .....	11
<b>Gambar 4.1</b> Grafik distribusi gaussian untuk nilai atribut R.....	17
<b>Gambar 4.2</b> Grafik distribusi gaussian untuk nilai atribut G.....	18
<b>Gambar 4.3</b> Grafik distribusi gaussian untuk nilai atribut B.....	18

## DAFTAR TABEL

	Halaman
<b>Tabel 3.1</b>	Model <i>confusion matrix</i> ..... 9
<b>Tabel 4.1</b>	Nilai atribut dataset citra apel <i>golden</i> ..... 12
<b>Tabel 4.2</b>	Nilai <i>centroid</i> awal untuk setiap variasi..... 13
<b>Tabel 4.3</b>	Pengelompokkan data untuk setiap variasi ..... 14
<b>Tabel 4.4</b>	Hasil akhir pengelompokkan data pada setiap variasi ..... 14
<b>Tabel 4.5</b>	<i>Confusion matrix</i> variasi A dengan akurasi sebesar 95,33% ..... 16
<b>Tabel 4.6</b>	<i>Confusion matrix</i> variasi B dengan akurasi sebesar 66,67% ..... 16
<b>Tabel 4.7</b>	<i>Confusion matrix</i> variasi C dengan akurasi sebesar 66,67% ..... 16
<b>Tabel 4.8</b>	<i>Confusion matrix</i> variasi D-G dengan akurasi sebesar 95,33% ..... 17

## DAFTAR LAMPIRAN

	Halaman
<b>Lampiran 1</b> Nilai atribut dataset citra apel <i>golden</i> .....	24
<b>Lampiran 2</b> Pengelompokkan akhir data setiap variasi .....	26
<b>Lampiran 3</b> Sintaks pengolahan citra menggunakan octave .....	29
<b>Lampiran 4</b> Sintaks simulasi menggunakan scilab .....	29

# BAB I

## PENDAHULUAN

### I.1 Latar Belakang

Indonesia memiliki sumber daya alam yang melimpah, terutama pada hasil perkebunannya seperti buah-buahan merupakan salah satu yang memiliki potensi besar. Hal ini tercermin dari keanekaragaman buah-buahan tropis di Indonesia. Apel adalah salah satu jenis buah dengan beragam varietas dan warna, diantaranya yaitu merah, hijau ataupun kuning. Dari banyaknya jenis buah apel tersebut tentu akan sedikit menyulitkan untuk membedakan antara buah apel yang satu dengan lainnya [1, 2].

Produksi pengolahan apel memerlukan proses *grading* berdasarkan tingkat kematangan, kualitas, kondisi buah maupun ukuran buah yang sesuai. Parameter kualitas biasanya ditentukan berdasarkan ukuran, berat, karakteristik warna dan bentuk buah. Warna dan tekstur adalah aspek utama gambar dan merupakan bagian penting dari pemahaman visual. Hal ini menjadi tantangan tersendiri pada proses sortir skala besar karena membutuhkan tenaga kerja yang banyak. Penyortiran varietas apel yang dilakukan menggunakan tenaga kerja manual tidak dapat diandalkan karena menimbulkan biaya tinggi, kebosanan, tidak konsisten dan tingkat akurasi yang rendah [3-8].

Untuk mengatasi masalah tersebut, diperlukan teknik yang dapat menjamin konsistensi kualitas tingkat kematangan buah apel dan sesuai dengan standar keamanan pangan sehingga keluhan konsumen dapat diminimalisir. Identifikasi yang akurat dari varietas apel dapat mendorong perkembangan yang sehat dan stabil dari industri apel global. Nilai apel bergantung pada kualitas buahnya. Oleh karena itu dibutuhkan suatu metode atau teknik yang dapat diandalkan untuk membedakan varietas dengan cepat dan tidak merusak apel [6-10].

Perkembangan teknologi yang semakin canggih memungkinkan identifikasi kualitas buah dan mengenali jenis buah apel dengan bantuan komputer [4]. Namun, identifikasi kualitas buah secara otomatis menggunakan komputer masih

merupakan tugas menantang karena terdapat beragam jenis buah-buahan khususnya buah apel [1, 11].

Berdasarkan hal yang telah dijelaskan diatas, maka akan dilakukan penelitian mengenai identifikasi kualitas buah apel khususnya apel *golden* melalui citra visual (aspek warna). Pada penelitian ini, akan digunakan algoritma *machine learning* dengan metode *unsupervised learning* yaitu *k-means clustering*. Penelitian ini diharapkan mampu membuat program *machine learning* yang dapat mengidentifikasi kualitas buah apel *golden* ke dalam tiga *cluster*, yaitu apel *golden 1*, apel *golden 2* dan apel *golden 3*.

## **I.2 Rumusan Masalah**

1. Bagaimana mengidentifikasi apel *golden* berdasarkan *dataset kaggle* dengan metode *k-means clustering*?
2. Bagaimana tingkat akurasi dari penerapan metode *k-means clustering* dalam mengidentifikasi kualitas apel *golden* berdasarkan warna citra?

## **I.3 Tujuan Penelitian**

1. Menghasilkan program *machine learning* yang mampu mengidentifikasi kualitas apel *golden* menggunakan metode *k-means clustering*.
2. Menganalisis tingkat akurasi dari penerapan metode *k-means clustering* dalam mengidentifikasi kualitas apel *golden* berdasarkan warna citra.

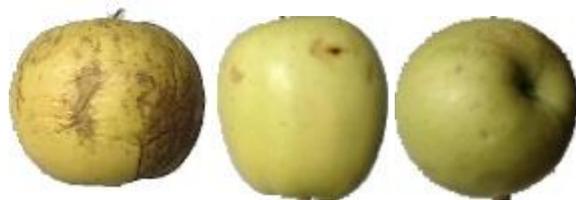
## BAB II

### TINJAUAN PUSTAKA

#### II.1 Apel

Buah apel adalah produk pertanian yang sangat populer dengan nilai gizi tinggi. Apel pertama kali dibudidayakan di wilayah Asia Tengah dan kemudian diperluas ke daerah yang lebih dingin. Apel yang dibudidayakan adalah kompleks hibrida interspesifik poliploid heterozigot. Apel telah dibudidayakan di Asia dan Eropa selama ribuan tahun dan merupakan spesies yang paling banyak ditanam dari *genus malus* karena kemampuan beradaptasi terhadap lingkungan yang tinggi. Produksi apel global mencapai 70 juta ton pada 2018-2019 [3, 9, 12].

Apel merupakan komoditas pertanian penting di pasar makanan segar secara global. Kualitas apel ditentukan oleh karakteristik eksternal seperti warna, ukuran dan tekstur permukaan, dan parameter internal seperti rasa manis, keasaman, kekencangan, tekstur jaringan, asam askorbat dan senyawa polifenol. Apel *golden* adalah varietas apel kuning yang populer, menurut sebuah laporan dari asosiasi apel *United State*. Apel *golden* menempati peringkat di bawah apel paling populer [8, 16].



**Gambar 2.1** Contoh citra apel *golden*

Nilai apel bergantung pada kualitas buahnya. Apel yang berkualitas tinggi memiliki harga jual yang tinggi pula, begitupun sebaliknya. Apel *golden* dibagi ke dalam tiga kelas yaitu apel *golden 1*, apel *golden 2* dan apel *golden 3*. Apel *golden 1* memiliki karakteristik warna kuning kecoklatan dengan permukaan agak kasar yang menandakan apel sudah matang. Apel *golden 2* memiliki karakteristik warna kuning dengan sedikit bercak kecoklatan yang menandakan apel mulai matang. Sedangkan apel *golden 3* memiliki karakteristik warna kuning kehijauan yang menandakan apel masih muda.

## II.2 Pengolahan Citra

Citra digital merupakan salah satu media yang paling penting untuk mengkomunikasikan informasi. Citra digital sangat efektif dalam memberi informasi atribut yang membantu dalam evaluasi tertentu. Mengekstrak dan memahami informasi dari gambar sehingga dapat digunakan untuk banyak tugas adalah fungsi utama *machine learning* [10, 17].

Pengolahan citra atau *image processing* adalah suatu sistem yang mengambil masukan (*input*) berupa citra (*image*) dan keluarannya (*output*) juga berupa citra (*image*). Awalnya pengolahan citra dilakukan hanya untuk meningkatkan kualitas citra. Namun seiring berkembangnya dunia komputasi yang ditandai dengan peningkatan kapasitas dan kecepatan pemrosesan komputer, serta munculnya ilmu-ilmu komputer yang memungkinkan manusia dapat mengambil informasi dari suatu citra maka pengolahan citra tidak dapat dilepaskan dengan bidang *computer vision* [18].

Pengolahan citra digital (*digital image processing*) adalah cabang ilmu yang mempelajari teknik pengolahan citra. Citra yang dimaksud adalah gambar ataupun video. Arti kata digital menunjukkan bahwa pengolahan citra dilakukan secara digital menggunakan komputer. Teknologi saat ini memungkinkan untuk melakukan klasifikasi dan identifikasi citra digital [1, 18].

Sebuah citra digital dapat direpresentasikan dengan matriks dua dimensi  $f(x,y)$  dengan  $M$  kolom dan  $N$  baris. Piksel atau elemen gambar merupakan elemen terkecil dari suatu citra yang terletak di persimpangan kolom dan baris dari metrik tersebut [18].

$$f(x,y) \approx \begin{matrix} f(0,0) & f(0,1) & \dots & f(0,M-1) \\ f(1,0) & f(1,1) & \dots & f(1,M-1) \\ \dots & \dots & \dots & \dots \\ f(N-1,0) & f(N-1,1) & \dots & f(N-1,M-1) \end{matrix}$$

Suatu citra  $f(x,y)$  dalam fungsi matematis memiliki batas fungsi yang dapat dituliskan sebagai berikut:

$$0 \leq x \leq M - 1$$

$$0 \leq y \leq N - 1$$

$$0 \leq f(x,y) \leq G - 1$$

dimana:

M : jumlah piksel baris pada *array* citra

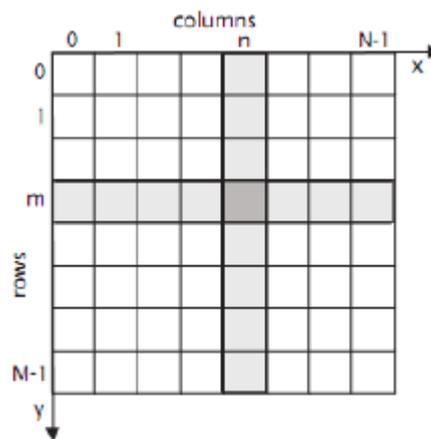
N : jumlah piksel kolom pada *array* citra

G : nilai skala keabuan (*graylevel*)

Secara umum, besarnya nilai M, N dan G merupakan perpangkatan dari bilangan biner.

$$M = 2^m; N = 2^n; G = 2^k$$

Dimana nilai m, n dan k merupakan bilangan bulat positif. Interval (0,G) merupakan skala keabuan (*grayscale*). Besar G bergantung pada proses digitalisasi. Pada umumnya, keabuan 0 menyatakan intensitas hitam dan 1 menyatakan intensitas putih.



**Gambar 2.2** Representasi citra digital dalam 2D [18].

Pengolahan citra digital memungkinkan untuk mengenali objek tertentu. Salah satu metode yang digunakan yaitu segmentasi warna. Normalisasi RGB merupakan salah satu metode segmentasi warna yang memiliki keunggulan yaitu mudah, cepat dan efektif pada objek rambu lalu lintas [18].

### II.3 Ruang Warna

Ruang warna atau model warna merupakan metode untuk menentukan, membuat, ataupun memvisualisasikan warna. Citra terdiri dari tiga warna penting yaitu merah, hijau dan biru. RGB (*Red, Green and Blue*) adalah model warna yang paling umum dalam pemrosesan citra. RGB atau juga disebut sebagai gambar *truecolor* yang mendefinisikan komponen warna merah, hijau dan biru

untuk setiap piksel individu. Pada dasarnya, untuk fitur warna setiap gambar masing-masing dipisahkan menjadi bidang merah, hijau dan biru. Lalu menghitung *mean*, median dan standar deviasi berdasarkan nilai RGB. *Array* RGB merupakan kelas ganda yang setiap komponen warna adalah nilai antara 0 sampai 1 [11, 15, 18, 19].

## **II.4 *Machine Learning***

*Machine learning* memiliki beragam algoritma, namun secara garis besar *machine learning* dibagi menjadi 2 bagian:

### **II.4.1 *Supervised Learning***

Model pembelajaran dibangun untuk membuat prediksi berdasarkan contoh *input* yang diberikan. Algoritma *supervised learning* mengambil dataset *input* untuk mempelajari model regresi/klasifikasi. Selanjutnya, algoritma ini melatih model untuk menghasilkan prediksi sebagai respon terhadap dataset uji [13].

### **II.4.2 *Unsupervised Learning***

Pelatihan algoritma *unsupervised learning* menggunakan data yang tidak diberi label dan memungkinkan algoritma untuk bertindak sendiri tanpa panduan. Sistem *unsupervised learning* dapat mengelompokkan data yang tidak disortir menurut persamaan dan perbedaan meskipun tidak ada kategori yang disediakan sebelumnya [13].

## **II.5 *Clustering***

*Clustering* pada dasarnya merupakan teknik pengklasifikasian atau mengelompokkan objek menjadi beberapa kelompok berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya [20-21]. *Clustering* adalah alat analisis data yang memecahkan masalah pengelompokkan. *Clustering* dilakukan dengan menggunakan beberapa pengukuran perhitungan jarak. Tugas komputasi membagi kumpulan piksel menjadi  $k$  himpunan bagian atau juga sering disebut pembelajaran tanpa pengawasan (*unsupervised learning*) [17-18].

## **II.6 *K-Means Clustering***

Beragam pendekatan *clustering* yang dirancang untuk berbagai tujuan, salah satu yang khas yaitu *k-means clustering* [10]. *K-means clustering* merupakan metode pengelompokkan objek yang berusaha mempartisi data menjadi beberapa

*cluster* sehingga data dalam satu kelompok mempunyai karakteristik berbeda dengan data yang ada pada kelompok lain. Simbol  $k$  menunjukkan konstanta jumlah *cluster* yang diinginkan, dan *means* berarti nilai rata-rata suatu group data yang dalam hal ini didefinisikan sebagai cluster [20-23].

*K-means clustering* merupakan metode *cluster* berbasis jarak yang melakukan partisi dataset ke dalam sejumlah *cluster* yang sudah ditentukan di awal. *K-means clustering* termasuk metode *unsupervised clustering* yang mengklasifikasikan objek data input ke dalam beberapa kelompok berdasarkan jarak inherennya satu sama lain dan algoritma ini hanya bekerja pada atribut numerik. Algoritma *k-means clustering* pada dasarnya sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi dan umum penggunaannya dalam praktek [10, 21, 23-25].

Proses dasar algoritma *k-means clustering* [18]:

1. Tentukan  $k$  sebagai jumlah *cluster* yang ingin dibentuk dan tetapkan *centroid* awal masing-masing *cluster*.
2. Hitung jarak setiap data ke *centroid* untuk masing-masing *cluster* ke- $k$  menggunakan persamaan Euclidean.

$$d_k = \sqrt{\sum_{i=1}^n (x_i - c_k)^2} \quad (2.1)$$

dimana:

$d_k$  = Jarak Euclidean setiap data terhadap *centroid cluster* ke- $k$ .

$n$  = Jumlah atribut

$x_i$  = Nilai atribut data ke- $i$

$c_k$  = *Centroid* pada *cluster* ke- $k$

3. Kelompokan data ke dalam *cluster* dengan jarak yang paling pendek menggunakan.
4. Hitung *centroid* yang baru menggunakan persamaan:

$$c_k = \frac{\sum_{i=1}^p y_i}{p} \quad (2.2)$$

dimana:

$c_k$  = Nilai *centroid* pada *cluster* ke- $k$

$p$  = Jumlah data yang ada dalam satu *cluster*

$y_i$  = Nilai atribut data ke- $i$  pada *cluster* ke- $k$