



## Kompresi Citra Digital Dengan Basis Komponen Warna RGB Menggunakan Metode *K-Means Clustering*

Arief Bramanto Wicaksono Putra<sup>1</sup>, Muhammad Trisna Aryuna<sup>2</sup> dan Rheo Malani<sup>3</sup>

<sup>1</sup>Jurusan Teknologi Informasi, Politeknik Negeri Samarinda, email : ariefbram@gmail.com

<sup>2</sup>Jurusan Teknologi Informasi, Politeknik Negeri Samarinda, email : trisnaaryuna21@gmail.com

<sup>3</sup>Jurusan Teknologi Informasi, Politeknik Negeri Samarinda, email : anaogie@gmail.com

### [1] Abstrak

*Dengan semakin berkembangnya teknologi dan media digital, kualitas data yang digunakan juga semakin tinggi namun ukuran dari data juga semakin besar dan membutuhkan media penyimpanan yang lebih besar. Kompresi data adalah solusi yang ditawarkan dengan tujuan mampu menghasilkan data dengan ukuran yang lebih kecil serta masih memiliki kualitas yang masih layak dan dapat di uji performansi. Pada penelitian ini metode k-means clustering akan digunakan untuk mengkompresi suatu data berupa citra digital. Dengan cara mengelompokkan warna suatu citra dan mengubah nilai piksel warna yang ada di dalam citra tersebut berdasarkan nilai pusat cluster pada masing-masing anggota cluster. Nilai centroid awal yang ditentukan pada tahap awal clustering akan berpengaruh terhadap hasil kompresi. Pada penelitian ini dilakukan percobaan sebanyak 10 kali, dengan hasil kualitas citra terbaik didapat pada percobaan ke 5 dengan nilai MSE sebesar 70,22 dan nilai PSNR sebesar 29,70. Sedangkan kualitas kompresi terbaik didapatkan pada percobaan ke 7 dengan rasio kompresi sebesar 74,5%. Didapatkan juga hasil pengukuran kualitas citra terendah terdapat pada percobaan ke 10 dengan nilai MSE sebesar 73,45 dan nilai PSNR sebesar 29,51, serta kualitas kompresi terendah didapatkan pada percobaan ke 3 dengan hasil rasio kompresi sebesar 71,3%. Rata-rata hasil pengukuran mendapatkan nilai MSE sebesar 71,47, nilai PSNR sebesar 29,62 dan rasio kompresi sebesar 72,40%.*

**Kata kunci:** Citra Digital, Kompresi, K-Means Clustering

### [2] Abstract

*With the development of technology and digital media, the quality of the data used is also getting higher but the size of the data is also getting bigger and requires larger storage media. To overcome the increasing need for data storage, one way that can be used is by compressing data to save space in storage memory. In this study, the k-means clustering method will be used to compress data in the form of a digital image. By grouping the colors of an image and changing the value of the color pixels in the image based on the value of the cluster center of each cluster member. The initial centroid value which is determined at the initial stage of clustering will affect the compression results. In this study, 10 experiments were carried out, with the best image quality results obtained in the 5th experiment with an MSE value of 70.22 and a PSNR value of 29.70. While the compression quality was obtained in the 7th experiment with a compression ratio of 74.5%. The results of the measurement of image quality in the 10th experiment were also obtained with an MSE value of 73.45 and a PSNR value of 29.51, and the lowest compression quality was obtained in the third experiment with a compression yield ratio of 71.3%. The average measurement results obtained an MSE value of 71.47, a PSNR value of 29.62 and a compression ratio of 72.40%.*

**Keywords:** Digital Image, Compression, K-Means Clustering

## 1. PENDAHULUAN

Dengan semakin berkembangnya teknologi informasi dan semakin mudahnya akses terhadap media digital. Maka berdampak pula terhadap kebutuhan penyimpanan data. Salah satu cara untuk menghemat penyimpanan adalah dengan mengompresi suatu data, baik itu data suara, teks, maupun citra agar dapat menghemat tempat pada penyimpanan. Proses kompresi merupakan proses mereduksi ukuran suatu data untuk menghasilkan representasi digital yang padat atau mampat (*compact*) namun tetap dapat mewakili kuantitas informasi yang terkandung pada data tersebut [1]. Pada penelitian ini akan dilakukan kompresi suatu data, data yang digunakan merupakan data citra digital.

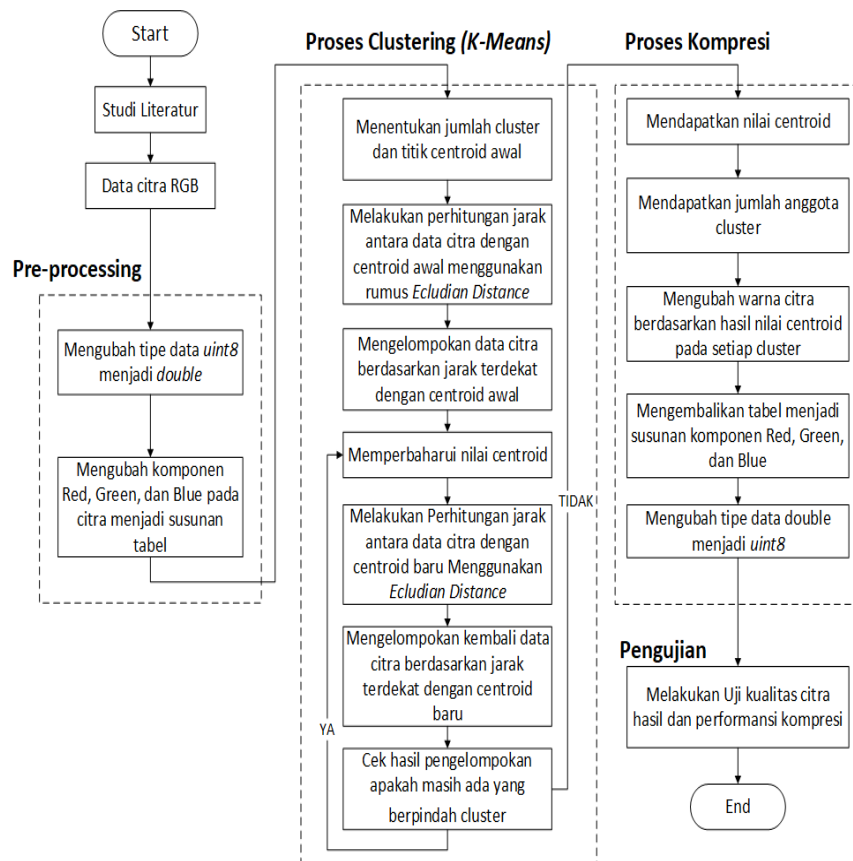
Kompresi citra digital bertujuan untuk mengurangi redundansi dari data-data yang terdapat dalam citra sehingga dapat disimpan atau ditransmisikan secara efisien [2]. Ada dua teknik mendasar dalam melakukan kompresi citra, yaitu teknik kompresi citra *lossless* dan *lossy* [3]. Teknik kompresi *lossless* merupakan teknik yang mengandalkan keutuhan informasi setelah dilakukan konstuksi citra, namun kelemahannya yaitu rasio kompresi citra teknik ini sangat rendah. Sedangkan teknik kompresi *lossy* merupakan teknik yang mengutamakan rasio kompresi sangat tinggi dengan mengorbankan beberapa informasi yang hilang, tetapi masih bisa ditolerir oleh persepsi mata manusia [4].

*Clustering* adalah sebuah proses untuk mengelompokkan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum [5], [6]. *K-Means* merupakan salah satu metode data *clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. [7]–[9]. Citra yang akan dikompresi adalah citra RGB dengan kedalaman 8 bit *color map*, hal ini yang menjadi dasar acuan dengan mengumpulkan kemiripan nilai pixel dalam setiap cluster akan memperkecil variasi nilai pixel yang berpengaruh terhadap ukuran data citra

Pada penelitian ini, proses kompresi citra akan dilakukan dengan menggunakan metode *k-mean clustering*. Dimana piksel warna pada suatu citra akan dikelompokkan terlebih dahulu berdasarkan *cluster* yang telah ditetapkan, sehingga didapatkan informasi tentang keanggotaan warna yang ada didalam citra. Setelah mendapatkan kelompok warna suatu citra dari hasil *clustering*, nilai piksel warna yang ada didalam citra tersebut akan diubah berdasarkan nilai titik pusat *cluster* (nilai *centroid*) pada masing-masing anggota *cluster*. Maka warna didalam citra tersebut yang semula terdiri dari berbagai macam warna akan menjadi sejumlah warna berdasarkan *cluster* yang ditetapkan.

## 2. METODE PENELITIAN

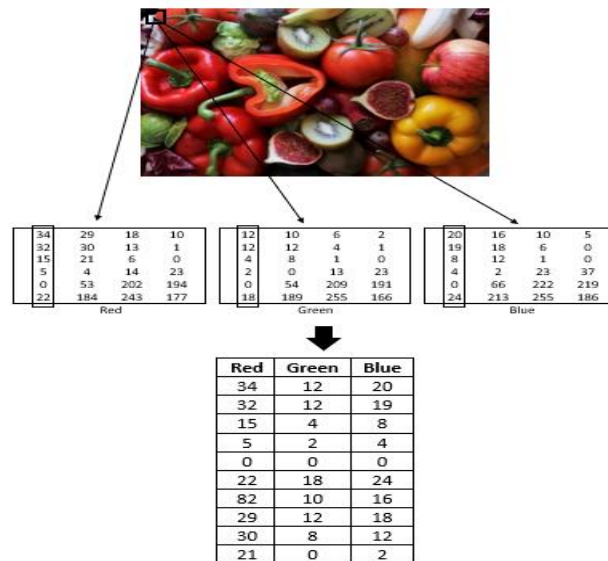
Dalam penelitian ini objek yang digunakan merupakan data citra berjenis citra *true color* dengan resolusi 100 x 100 piksel. Dalam *pre-processing*, objek akan dilakukan perubahan susunan citra hingga menghasilkan data siap olah. Setelah tahap *pre-processing*, data akan dilakukan pengelompokan dengan menggunakan metode *k-means* hingga menghasilkan kelompok-kelompok warna beserta anggotanya. Jumlah kelompok warna akan digunakan sebagai dasar untuk mengubah piksel citra yang semula terdiri dari berbagai macam warna akan menjadi beberapa warna berdasarkan jumlah kelompok yang ditetapkan. Adapun tahapan-tahapan penelitian yang dilakukan untuk kompresi citra dengan menggunakan metode *k-means clustering* disajikan pada Gambar 1.



Gambar 1. Metode Penelitian

a. Pre-Processing

Pada tahap *pre-processing*, data citra yang telah terbaca pada program pengolahan citra akan diubah tipe datanya. Tipe data yang awalnya *uint8* diubah menjadi tipe data *double* agar dapat dilakukan proses aritmetika pada komponen citra. Data citra dengan komponen *red*, *green*, dan *blue* akan disusun secara kolom per kolom menjadi satu kolom hingga menjadi data siap olah. Contoh dari penyusunan komponen citra dapat dilihat pada Gambar 2.



Gambar 2. Perubahan susunan komponen citra

### b. Clustering

Setelah mendapatkan tabel data citra, selanjutnya data akan dilakukan proses pengelompokan dengan menggunakan algoritma *k-means clustering*. Algoritma *k-means clustering* merupakan sebuah algoritma yang mengelompokkan data berdasarkan jarak terdekat dari suatu *cluster*. Proses *clustering* pada penelitian ini digunakan untuk mengelompokkan warna pada citra. Jumlah *cluster* yang digunakan adalah sebanyak 8 *cluster* dan akan dilakukan percobaan sebanyak 10 kali dengan iterasi maksimal sebanyak 100 kali. Pembentukan nilai *centroid* awal diambil berdasarkan nilai piksel hasil pencampuran tingkat intensitas warna pokok *red*, *green*, dan *blue*. Setelah *centroid* awal terbentuk dan jumlah *cluster* ditetapkan, selanjutnya melakukan perhitungan jarak antar data dengan *centroid* awal menggunakan rumus *euclidean distance* [10]. Rumus perhitungan jarak dapat dilihat pada Persamaan 1.

$$d = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (1)$$

Keterangan :

$n$  = dimensi data.

$x_i$  = data nilai piksel citra untuk  $k = 1, 2, \dots, n$

$x_j$  = data nilai centroid untuk  $k = 1, 2, \dots, n$

Hasil jarak *euclidean* terkecil dari *centroid* yang akan diambil sebagai dasar untuk mengelompokkan data citra. Lalu hitung kembali data citra dengan keanggotaan *cluster* yang telah didapat untuk mendapatkan nilai pusat *cluster* baru. Pusat *cluster* baru adalah rata-rata dari semua data/objek dalam *cluster* tertentu yang dapat dihitung dengan persamaan 2.

$$C = \frac{\sum m}{n} \quad (2)$$

Keterangan :

$C$  = Titik pusat *cluster/centroid* data.

$m$  = Jumlah nilai data pada anggota *cluster*.

$n$  = Jumlah anggota *cluster*.

Apabila pada saat proses perhitungan jarak terjadi adanya pergeseran anggota *cluster*, maka perlu membangkitkan nilai *centroid* baru pada iterasi berikutnya. Setelah memperoleh *centroid* baru, langkah selanjutnya antara lain melakukan perhitungan *euclidean distance*, proses pengelompokan berdasarkan jarak terdekat, pengecekan perubahan anggota *cluster*, hingga proses pembangkitan *centroid* baru yang akan dilakukan jika adanya perubahan anggota *cluster*. Langkah-langkah tersebut akan terus diulang hingga tidak adanya perubahan anggota *cluster*, dan proses iterasi dihentikan.

### c. Kompresi

Setelah proses *clustering* selesai, maka akan didapatkan hasil berupa jumlah keanggotaan *cluster* beserta nilai titik pusat dari *cluster* tersebut. Jumlah anggota *cluster* akan digunakan sebagai jumlah warna yang ada didalam piksel citra yang semula terdiri dari berbagai macam warna akan menjadi sejumlah nilai berdasarkan anggota *cluster* yang ditetapkan. Sedangkan nilai titik pusat *cluster (centroid)* akan digunakan sebagai nilai piksel pada citra dengan mengikuti keanggotaan *cluster* tersebut. Selanjutnya, susunan data yang berbentuk tabel akan dikembalikan menjadi susunan komponen *red*, *green*, dan *blue* dan mengubah kembali tipe data menjadi uint8 agar dapat ditampilkan kembali dalam bentuk citra.

#### d. Unjuk Kerja

Setelah mendapatkan hasil citra terkompresi, selanjutnya akan dilakukan pengujian dengan menggunakan *Mean Square Error* (MSE) dan *Peak Signal-to-Noise Ratio* (PSNR) sebagai parameter pengukuran kualitas citra hasil kompresi. *Mean Square Error* (MSE) adalah nilai error kuadrat rata-rata antara citra asli dengan citra hasil kompresi. MSE menunjukkan kuadrat rata-rata selisih nilai piksel citra rekonstruksi dengan citra sumber. Semakin kecil nilai MSE maka kualitas citra hasil kompresi juga baik. Untuk menghitung nilai MSE dapat menggunakan persamaan 3.

$$MSE = \frac{1}{M \cdot N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [f(i,j) - g(i,j)]^2 \quad (3)$$

Keterangan :

$i$  dan  $j$  = Koordinat dari piksel citra.

$M$  dan  $N$  = Dimensi dari citra.

$f(i,j)$  = Citra hasil kompresi.

$g(i,j)$  = Citra asli sebelum dikompres.

PSNR (*Peak Signal-to-Noise Ratio*) adalah perbandingan antara nilai maksimum dari sinyal yang diukur dengan besarnya derau yang berpengaruh pada sinyal tersebut. PSNR memiliki satuan decibel (db). PSNR digunakan untuk mengetahui kualitas citra sebelum dan sesudah kompresi. Nilai PSNR bergantung dengan nilai MSE. Semakin besar nilai PSNR, maka kualitas citra hasil kompresi juga semakin baik. Jika  $MSE = 0$ , maka  $PSNR = \infty$ . Untuk menghitung nilai PSNR dapat menggunakan persamaan 4.

$$PSNR = 10 \log_{10} \frac{255^2}{MSE} \quad (4)$$

Dimana 255 merupakan nilai maksimum dari suatu pixel dan MSE (*Mean Square Error*) merupakan nilai error kuadrat rata-rata dari citra asli dengan citra hasil kompresi. Untuk mengukur kinerja dari pemampatan yang telah dibuat dapat menggunakan rasio kompresi sebagai salah satu parameter pengukuran. Rasio kompresi ini bertujuan untuk membandingkan ukuran citra asli sebelum kompresi terhadap citra hasil dari proses kompresi. Adapun rumus rasio kompresi ditunjukkan pada persamaan 5.

$$Rasio = \frac{(uAsli - uKompresi)}{uAsli} \times 100\% \quad (5)$$

Keterangan :

$uAsli$  = Ukuran file citra asli.

$uKompresi$  = Ukuran file citra hasil kompresi.

### [3] HASIL DAN PEMBAHASAN

Material yang digunakan pada penelitian ini merupakan data citra yang diambil dari dataset image pada website pixabay.com yang bersifat *free license*. Data citra yang digunakan merupakan citra buah yang beresolusi 100 x 100 piksel dengan ukuran file 22.1 kilobyte (KB) dan dengan format citra PNG. Data citra yang digunakan disajikan pada Gambar 3.



Gambar 3. Data Citra

Setelah mengubah tipe data menjadi *double* maka selanjutnya dilakukan proses penyusunan komponen *red*, *green*, dan *blue* menjadi susunan tabel piksel. Penyusunan akan dilakukan seperti pada Gambar 2. dan menghasilkan data berjumlah 10.000 data piksel dengan label ciri yang dimiliki tiap piksel yaitu berupa komponen *red*, *green*, dan *blue*. Hasil dari perubahan disajikan pada Tabel 1.

Tabel 1. Data Piksel Citra

Data ke-	Red	Green	Blue
1	34	12	20
2	32	12	19
3	15	4	8
...	...	...	...
5001	205	184	113
5002	178	153	71
5003	167	140	58
...	...	...	...
9998	63	54	53
9999	69	58	56
10000	75	64	62

Pada penelitian ini, nilai *centroid* awal akan diambil berdasarkan hasil dari percobaan pencampuran tingkat intensitas warna pada kanal *red*, *green*, dan *blue* secara random. Besaran random yang digunakan dapat dilihat pada Tabel 2.

Tabel 2. Tingkat Intensitas Piksel

Percobaan	Tingkat Intensitas		
	R	G	B
1	0 – 30 %	0 – 30 %	0 – 30 %
2	31 – 79 %	31 – 79 %	31 – 79 %
3	80 – 100 %	80 – 100 %	80 – 100 %
4	50 – 100 %	0 – 50 %	0 – 50 %
5	0 – 50 %	50 – 100 %	0 – 50 %
6	0 – 50 %	0 – 50 %	50 – 100 %
7	50 – 100 %	50 – 100 %	0 – 100 %
8	50 – 100 %	0 – 100 %	50 – 100 %
9	0 – 100 %	50 – 100 %	50 – 100 %
10	0 – 100 %	0 – 100 %	0 – 100 %



<b>9998</b>	63	54	53	2	1	7	4	3	6	3	7	1	8
<b>9999</b>	69	58	56	2	1	7	4	3	6	3	7	1	3
<b>10000</b>	75	64	62	2	1	7	4	3	6	3	7	1	3

Sedangkan hasil nilai *centroid* dari proses *clustering* pada iterasi terakhir percobaan 1 hingga 5 disajikan pada Tabel 6 dan iterasi terakhir percobaan 6 hingga 10 pada Tabel 7.

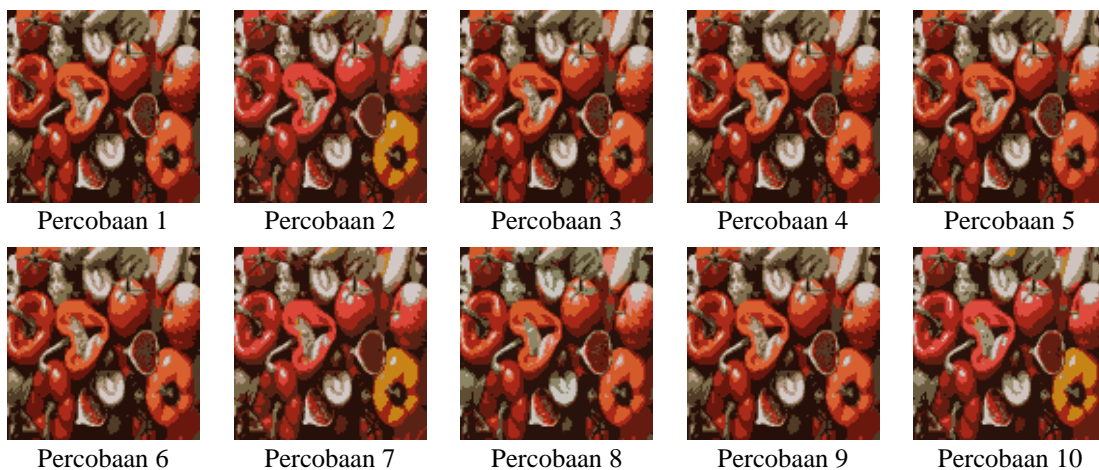
Tabel 6. Nilai Centroid Hasil Iterasi Terakhir Percobaan 1-5

C	P1			P2			P3			P4			P5		
	R	G	B	R	G	B	R	G	B	R	G	B	R	G	B
1	129	113	77	91	34	21	106	23	13	105	23	13	171	43	26
2	81	60	39	38	16	9	39	15	9	216	94	46	39	15	9
3	171	43	26	158	37	22	129	113	78	39	15	9	81	60	39
4	39	15	9	175	149	125	171	43	26	81	60	39	106	23	13
5	181	152	130	209	199	196	216	96	47	129	112	77	210	202	199
6	106	23	13	118	102	70	181	152	131	181	151	128	129	113	78
7	210	201	199	219	70	56	81	60	39	209	201	198	181	152	130
8	216	96	47	199	131	21	210	202	199	170	42	26	216	95	47

Tabel 7. Nilai Centroid Hasil Iterasi Terakhir Percobaan 6-10

C	P6			P7			P8			P9			P10		
	R	G	B	R	G	B	R	G	B	R	G	B	R	G	B
1	129	113	77	157	37	22	101	27	15	81	59	39	166	39	24
2	216	96	47	219	70	56	203	197	192	106	23	13	160	144	115
3	106	23	13	90	34	21	40	18	11	216	95	47	105	91	61
4	171	43	26	199	131	21	211	91	42	39	15	9	199	132	22
5	181	152	130	175	148	125	219	134	115	181	151	129	40	18	11
6	81	60	39	38	16	9	168	39	25	171	43	26	220	75	62
7	39	15	9	209	199	196	95	78	51	209	201	198	208	195	191
8	210	201	199	118	102	70	141	134	101	129	112	77	99	29	17

Setelah mendapatkan nilai *centroid* dan anggota pada tiap *cluster*, tahapan berikutnya akan dilakukasn proses kompresi. Proses yang dilakukan yaitu data citra awal akan diubah berdasarkan nilai titik pusat *cluster (centroid)* pada masing-masing anggota *cluster* yang telah didapatkan pada proses *clustering*. Sehingga nilai piksel yang telah terkelompok pada tiap *cluster* akan diubah dengan nilai titik pusat *clusternya*. Data komponen citra yang telah diubah nilai pikselnya berdasarkan nilai *centroid* masih dalam susunan tabel dengan jumlah 10.000 data. Selanjutnya akan dilakukan proses pengembalian menjadi susunan piksel citra dengan ukuran 100 x 100 piksel dan mengubah tipe data *double* menjadi *uint8* agar citra dapat divisualisasikan. Berikut merupakan hasil dari tahapan kompresi disajikan pada Gambar 4.





**Gambar 4. Hasil Kompresi**

Setelah mendapatkan hasil citra terkompresi, tahapan selanjutnya akan dilakukan pengujian citra hasil kompresi untuk mengukur kualitas citra dan kinerja dari proses kompresi yang telah dilakukan. Berikut merupakan hasil dari pengukuran citra dapat dilihat pada tabel 10.

**Tabel 8. Hasil Pengukuran Performa dan Kualitas Kompresi**

Percobaan	Ukuran Citra (KB)		MSE	PSNR	Ukuran Kompresi (KB)	Rasio Kompresi
	Awal	Hasil				
1	22.125	6.318	70.450	29.686	15.807	71.44%
2	22.125	5.750	73.086	29.527	16.375	74.01%
3	22.125	6.359	70.276	29.697	15.766	71.26%
4	22.125	6.305	70.275	29.697	15.820	71.50%
5	22.125	6.318	70.217	29.700	15.807	71.44%
6	22.125	6.318	70.450	29.686	15.807	71.44%
7	22.125	5.634	73.276	29.515	16.491	74.54%
8	22.125	6.003	72.794	29.544	16.122	72.87%
9	22.125	6.279	70.396	29.689	15.846	71.62%
10	22.125	5.788	73.446	29.505	16.337	73.84%

#### [4] KESIMPULAN

Dengan melihat hasil penelitian yang telah dilakukan, maka dapat ditarik kesimpulan sebagai berikut, bahwa metode *k-means clustering* dapat digunakan untuk melakukan kompresi citra digital dan dapat berjalan dengan baik. Nilai *centroid* awal yang ditentukan pada tahap awal *clustering* akan berpengaruh terhadap hasil kompresi. Dengan dilakukan percobaan sebanyak 10 kali, telah didapatkan hasil kualitas citra yang terbaik pada percobaan ke 5 dengan nilai MSE sebesar 70,22 dan nilai PSNR sebesar 29,70. Untuk kualitas kompresi terbaik didapatkan pada percobaan ke 7 dengan rasio kompresi sebanyak 74,5% dari citra asli yang berarti semakin besar nilai rasio kompresi maka ukuran file citra akan semakin kecil. Namun juga memiliki kekurangan yaitu semakin besar citra terkompresi maka kualitas citra juga akan semakin menurun. Sedangkan pada percobaan ke 10 mendapatkan hasil pengukuran kualitas citra terendah dengan nilai MSE sebesar 73,45 dan nilai PSNR sebesar 29,51 serta pada percobaan ke 3 didapatkan hasil rasio kompresi terendah dengan menghasilkan rasio 71,3% dari citra asli, yang berarti citra pada percobaan ini paling sedikit kehilangan informasi ketika dikompresi dibandingkan dengan pada percobaan lain. Untuk rata-rata hasil dari semua percobaan didapatkan nilai MSE sebesar 71,47 , nilai PSNR sebesar 29,62 dan rasio kompresi sebesar 72,40%.

#### DAFTAR PUSTAKA

- [1] D. Putra, "Pengolahan Citra Digital," ANDI, Yogyakarta, 2010, ISBN : 9789792914436
- [2] S. F. Florentin and M. A. Rony, "Penerapan Kompresi Citra Digital Dengan Metode Kuantisasi Warna Menggunakan Algoritma K-Means," Jurnal SKANIKA, vol. 1, No. 2, pp. 431–438, 2018.
- [3] I. G. N. J. Suryaningrat, G. D. Sanjaya, R. Hadi, and N. L. G. P. Suwirmayanti, "Kompresi Citra Digital Menggunakan Metode Discrete Cosine Transform," Jurnal SINTAK, Vol.2, No.1, pp. 38–44, 2018.
- [4] P. N. Andono, T. Sutojo, and Muljono, "Pengolahan Citra Digital", ANDI, Yogyakarta, 2018, ISBN : 9789792963700.
- [5] P.-N. Tan, M. Steinbach, and V. Kumar, "Introduction to Data Mining," Pearson,

- NewYork, 2019, ISBN: 978-0321321367.
- [6] I. W. A. Wijaya and A. Kusumadewi, "Penerapan Algoritma K-Means Pada Kompresi Adaptif Citra Medis Mri," *J. Inform.*, vol. 11, no. 2, pp. 139–151, 2016.
  - [7] A. M. Alhumaerah, A. B. W. Putra, and Rihartanto, "Penerapan K-Means Clustering dalam Pengelompokan Lipstik Matte Lip Cream Berdasarkan Warna RGB," *Jurnal JST*, vol. 5, no. 1, pp. 28-35, 2019.
  - [8] R. I. Armianti, A. F. O. Gaffar, and A. B. W. Putra, "Penerapan K-Means Clustering Untuk Seleksi Frame Dominan," *Jurnal JTIK*, vol.7, no.4, pp. 745–754, 2018.
  - [9] P. R. Nastiti and A. B. W. Putra, "Perbandingan Algoritma K-Means Dan Fuzzy C-Means Clustering Untuk Kualifikasi Data Kinerja Dosen Di Jurusan Teknologi Informasi Polnes," *Seminar Nasional SEBATIK*, vol. 1, no. 1, pp. 71–76, 2017.
  - [10] M. R. Rosada, A. Bramanto, W. Putra, and A. F.O Gaffar, "Prototype Region Of Interest Citra Wajah Manusia Berbasis Binary Large Object (BLOB) Analysis," *Jurnal Dinamika Rekayasa*, vol. 16, no. 2, pp. 1-8, 2020.