

## Klasterisasi Citre Liburan Secara Otomatis Berbasis Segmentasi Area Menggunakan Metode K-Means

Linggar Bagas Saputro<sup>1</sup>, Yufis Azhar<sup>2</sup>, Agus Eko Minarno<sup>3</sup>

Universitas Muhammadiyah Malang

e-mail: linggar.bagas10@gmail.com<sup>1</sup>, yufis.az@gmail.com<sup>2</sup>, agoes.minarno@gmail.com<sup>3</sup>

### Abstrak

Mengabadikan momen liburan merupakan hal yang biasa dilakukan wisatawan lokal atau wisatawan domestik. Pada kamera digital citra tersimpan dalam bentuk file yang disimpan pada memori. Dalam memori citra yang tersimpan tidak tertata dengan rapi. Banyak dan beragamnya lokasi atau momen yang diabadikan menyulitkan dalam pengelompokan citra. Dalam hal ini diperlukannya pengelompokan citra liburan untuk mempermudah dalam pengelompokan citra berdasarkan lokasi. Dataset yang digunakan merupakan dataset citra liburan berjumlah 5000 data dari 49 lokasi wisata. Algoritma yang digunakan dalam penelitian ini adalah K-means. K-means merupakan salah satu teknik clustering yang terkenal mudah dan mampu mengkluster data besar dengan cepat. Pada penelitian ini citra disegmentasi menjadi 5 area, segmentasi citra dilakukan sebelum pengelompokan dengan metode K-means. Segmentasi citra dimaksudkan untuk meningkatkan kualitas kluster. Silhouette Coefficient dan Purity Measure digunakan untuk mengevaluasi model yang diusulkan. Pada evaluasi yang dilakukan model yang diusulkan mampu mengkluster citra liburan dengan baik. Hal ini dapat dibuktikan dalam 4 kali pengujian 3 diantaranya model segmentasi 5 lebih unggul dibandingkan model tanpa segmentasi pada pengujian purity measure. Evaluasi tambahan dilakukan untuk mengetahui pengaruh ruang warna terhadap penelitian dengan membandingkan nilai purity dari ruang warna RGB dan HSV dan memberikan hasil kesimpulan bahwa ruang warna memiliki pengaruh yang tidak terlalu signifikan pada penelitian ini.

**Kata kunci:** k-means, segmentasi, liburan.

### Abstract

Capture the moments of the holiday is common for local tourists or domestic tourists. In a digital camera the image is stored in the form of a file that is stored in memory. In memory the stored images are not neatly organized. The many and varied locations or moments that are enshrined make it difficult to group images. In this case it is necessary to group holiday images to facilitate the grouping of images by location. The dataset used is a holiday image dataset totaling 5000 data from 49 tourist sites. The algorithm used in this study is K-means. K-means is a clustering technique that is famous for being easy and able to cluster large data quickly. In this study the image is segmented into 5 areas, image segmentation is done before grouping with the K-means method. Image segmentation is intended to improve cluster quality. Silhouette Coefficient and Purity Measure are used to evaluate the proposed model. In the evaluation conducted the proposed model is able to cluster vacation images well. This can be proven in 4 tests 3 of which 5 segmentation models are superior to models without segmentasi in the purity measure test. Additional evaluation was conducted to determine the effect of color space on research by comparing the purity values of the RGB and HSV color spaces and providing the conclusion that the color space has a not too significant effect in this study.

**Keywords:** k-means, segmentasi, holiday.

### 1. Pendahuluan

Pariwisata merupakan keseluruhan elemen-elemen terkait yang didalamnya terdiri dari wisatawan, daerah tujuan wisata, perjalanan, industri dan lain sebagainya yang merupakan kegiatan pariwisata[1]. Indonesia merupakan negara yang memiliki banyak objek pariwisata untuk dikunjungi, misalnya wisata alam, sosial maupun wisata budaya[1]. Tingginya minat masyarakat yang ingin berkunjung di berbagai objek pariwisata, membuat industri pariwisata

dapat mendatangkan manfaat dan keuntungan untuk sebagian besar masyarakat Indonesia dan sebagai sumber devisa negara[2][3]. Objek pariwisata yang menyuguhkan panorama yang indah membuat para wisatawan mengabadikan setiap momen diberbagai lokasi.

Beragam momen dan lokasi yang diabadikan tersimpan didalam memori dan tidak tertata rapi. Hal ini akan menyusahkan wisatawan ketika akan melihat dan mengelompokkan foto berdasarkan lokasi. Banyaknya jumlah foto, akan sulit dilakukan pengelompokan secara manual dan hal tersebut akan memakan waktu yang cukup lama. Berdasarkan permasalahan diatas, perlu adanya sebuah sistem yang dapat mengelompokkan citra secara otomatis berdasarkan lokasi foto. Dalam pengelompokan citra terdapat metode seperti pada penelitian yang dilakukan oleh Atriyon Trisnawan, Wahyudi Harianto, dan Syahminan menggunakan metode K-means. Penelitian tersebut menggunakan dataset jenis beras yang beredar di pasaran, yaitu beras pandan wangi, rojo lele, IR 64, IR42, C 4, dan lain-lain. Pengujian dilakukan untuk mengetahui nilai akurasi metode K-means. Sebelum dilakukan proses pengelompokan dataset dilakukan ekstraksi fitur warna RGB untuk mendapatkan nilai red, green, blue pada setiap citra. Berdasarkan hasil pengujian yang dilakukan sebanyak 40 kali terdapat kesalahan sebesar 9 kali, dan kesesuaian 31 kali. Sehingga dapat disimpulkan bahwa tingkat keakuratan sebesar 77,5%[4]. Terdapat penelitian lain menggunakan metode yang sama dengan menggunakan data citra sawah dari pesawat tidak berawak (UAV). Pertama mensegmentasi area gabah menjadi 4 area menggunakan tinggi rendah image RGB yang didapat dari UAV, kemudian dilakukan pemotongan grafik untuk mengekstraksi latar depan dan latar belakang gambar, gambar RGB latar depan dikonversi menjadi warna lab dan K-Means clustering digunakan untuk memberikan label pada piksel berdasarkan informasi warna. Luas butir padi dalam gambar dihitung dari gambar berkerumun. Menggunakan area biji-bijian ini, hasil panen dapat di perkirakan. Penelitian menunjukkan bahwa metode yang diusulkan dapat mengelompokkan area biji-bijian dengan kesalahan relatif 6%-33% untuk memperbaiki kesalahan relatif dari metode sebelumnya sebesar 1%-31%, kesalahan relatif untuk empat bagian lapangan pada penelitian adalah 21%-31%[5].

K-Means merupakan salah satu algoritma clustering . Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok. K-Means merupakan *partitioning clustering* yang memisahkan data ke k daerah bagian yang terpisah[6]. Kelebihan K-Means algorithm ini mudah dan mampu untuk mengklaster data besar dan data outlier dengan sangat cepat[7]. Dalam implementasi algoritma K-Means, penelitian ini menggunakan Bahasa Python. Mudahnya dan cocoknya Python sebagai Bahasa pemrograman pengolah data dapat mempermudah dalam melakukan penelitian. Dalam melakukan pengujian, penelitian ini akan dilakukan pengukuran performa dari sistem menggunakan silhouette coefficient dan purity measure. Silhouette coefficient merupakan pengujian untuk melihat kualitas dan kekuatan cluster, seberapa baik suatu objek ditempatkan dalam suatu cluster[8][9]. Purity measure merupakan pengujian untuk melihat nilai kemurnian pada suatu cluster[10]. Dari hasil penelitian ini diharapkan tingkat keakurasian yang baik.

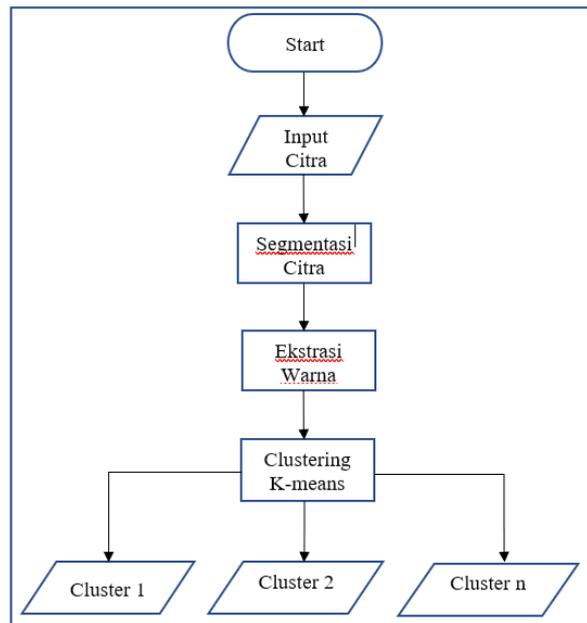
Segmentasi citra adalah proses untuk mempartisi atau membagi citra menjadi beberapa segmen atau bagian. Segmentasi gambar biasanya digunakan untuk menemukan objek atau batasan pada suatu citra[11]. Seperti pada penelitian Kai Tian, Jiunhao Li, Jiefeng zeng, Asenso Evans, Lina Zhang, segmentasi diterapkan pada tanaman tomat dengan background putih untuk menemukan objek daun[6]. Terdapat penelitian lain yang menggunakan segmentasi yang sama, dilakukan pada objek buah apel, dimana citra buah apel pada pohon disegmentasi untuk memisahkan objek buah apel dengan latar belakangnya[12].

Berdasarkan kelebihan dari algoritma K-Means, adapun hipotesa dari penelitian ini adalah sistem klasterisasi citra liburan secara otomatis berbasis segmentasi area menggunakan metode K-Means dengan Bahasa Python ini mampu menghasilkan sistem dengan nilai silhouette dan purity yang baik.

## 2. Metode Penelitian

Dataset Sistem clustering citra liburan ini merupakan sebuah perangkat lunak yang digunakan untuk mengelompokkan citra liburan. Sebelum dilakukan pengklasteran terhadap citra, pertama citra dirubah kedalam ruang warna RGB atau HSV, kemudian citra disegmen menjadi 5 bagian hal ini bertujuan untuk membagi citra menjadi beberapa area. Citra yang telah disegmentasi di ekstraksi berdasarkan ruang warna RGB atau HSV yang kemudian hasil ekstraksi warna disimpan dalam bentuk file .csv. Hasil ekstaksi warna yang tersimpan dalam file .csv akan dimasukkan pada algoritma k-means untuk dilakukan pengklasteran. Hasil akhir dari

proses ini adalah nilai purity yang didapatkan dari proses pengujian. Berikut arsitektur dari sistem clustering citra liburan, dapat dilihat pada Gambar 1.



Gambar 1. Model sistem

## 2.1 Dataset

Dataset yang digunakan dalam penelitian ini yaitu data citra liburan yang berjumlah 5000 data dari 49 lokasi wisata. Berikut beberapa contoh dataset, dapat dilihat pada Gambar 2.

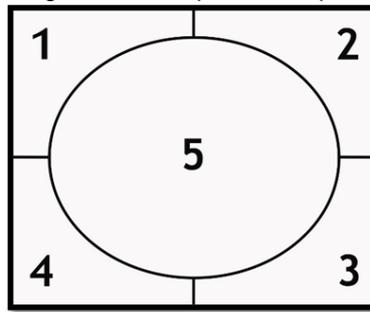


Gambar 2. Dataset Citra liburan

## 2.2 Segmentasi Area

Dalam visi komputer, segmentasi adalah proses mempartisi citra digital menjadi beberapa segmen (set piksel, juga dikenal sebagai super pixel). Tujuan dari segmentasi adalah untuk menyederhanakan dan / atau mengubah penyajian citra ke suatu yang lebih bermakna dan lebih mudah untuk menganalisis. Gambar segmentasi biasanya digunakan untuk menemukan objek dan batas-batas (garis, kurva, dll) dalam citra. Segmentasi citra adalah proses untuk menempatkan label untuk setiap pixel dalam sebuah citra, sehingga pixel dengan label yang sama dengan karakteristik visual tertentu. Hasil segmentasi citra adalah seperangkat segmen

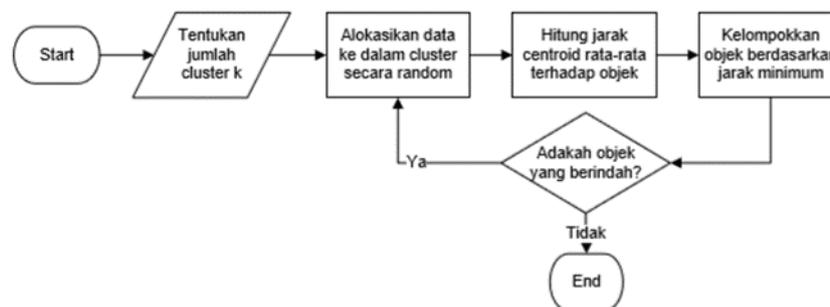
yang secara kolektif mencakup seluruh citra atau satu set kontur diekstrak dari citra (lihat deteksi tepi). Berikut merupakan proses segmentasi, dapat dilihat pada Gambar 3.



Gambar 3. Segmentasi Area

### 2.3 K-means

*K-Means* merupakan salah satu algoritma clustering. Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok. *K-Means* termasuk partitioning clustering yang memisahkan data ke  $k$  daerah bagian yang terpisah[16]. *K-Means* algorithm ini sangat terkenal karena mudah dan mempunya untuk mengklaster data besar dan data outlier dengan sangat cepat[9]. Sesuai dengan karakteristik partitioning clustering setiap data masuk ke dalam cluster tertentu. Memungkinkan bagi setiap data yang masuk dapat berpindah ke cluster yang lain pada tahapan berikutnya[15]. Berikut merupakan alur dari *K-Means* yang dapat dilihat pada Gambar 4.



Gambar 4. Algoritma K-measn

## 3. Hasil dan Pembahasan

Terdapat 3 pengujian yang dilakukan pada penelitian ini. Pengujian menggunakan jumlah data yang berbeda. Data yang digunakan 5000 citra liburan dan diukur dengan *silhouette coefficient*. Pengujian 2 dilakukan dengan menggunakan data lebih sedikit diukur menggunakan *purity measure*, dab pengujian 3 dilakukan dengan membandingkan ekstraksi HSV dan RGB.

### 3.1 Segmentasi

Pada proses ini dilakukan segmentasi terhadap citra menjadi 5 bagian. Bentuk segmentasi citra dapat dilihat pada Gambar 5.



Gambar 5. Bentuk Segmentasi

### 3.2 Pengujian 1

Pengujian 1 dilakukan terhadap citra yang disegmentasi menjadi 5 bagian dibandingkan dengan citra tanpa segmentasi, data yang digunakan merupakan data set pribadi 5000 citra yang diambil dari citra liburan di 49 lokasi. Pengujian ini dilakukan untuk menguji keakuratan k-means berdasarkan jarak rata-rata setiap objek dalam mengelompokkan citra, dimana hasil pengujian direpresentasikan dengan nilai silhouette coefficient. Berikut merupakan hasil dari pengujian 1 sekenario 1 yang dilakukan pada penelitian ini dapat dilihat pada Tabel 1.

*Tabel 1 Hasil Pengujian 1*

Uji	Data	Segmen	K	Silhouette
1	5000	5	49	0.07892
2	5000	none	49	0.13762

Tabel 1 menjelaskan hasil pengujian yang telah dilakukan pada skenario 1 pengujian 1. Hasil nilai silhouette coefficient pada data set 5000 dengan citra Nonsegmentasi memiliki nilai yang lebih tinggi daripada citra dengan Segmentasi 5. Hal ini bisa terjadi karena data yang dikelompokkan berjumlah besar dan memiliki varian warna yang relative sama, dan K-means merupakan metode yang bias mengelompokkan data dalam skala 2 dimensi. Silhouette coefficient adalah jenis evaluasi yang menghitung jarak rata-rata antar objek, sehingga ketika terdapat objek yang memiliki jarak rata-rata sama dengan 2 atau lebih objek, sangat memungkinkan objek itu akan tergabung dalam kluster yang salah sehingga mempengaruhi nilai dari silhouette coefficient. Dengan hasil dan kesimpulan diatas, diperlukan pengujian yang lebih lanjut dan mendalam pada penelitian ini untuk mengetahui lebih jelas keefektifan model ini.

### 3.2 Pengujian 2

Pada pengujian 2 dilakukan terhadap citra yang disegmentasi menjadi 5 bagian dibandingkan dengan citra tanpa disegmentasi. Data yang digunakan merupakan data citra liburan yang sama dengan pengujian 1, namun dengan menggunakan data yang lebih sedikit dengan 4 kali pengujian, dimana setiap pengujian dilakukan penambahan jumlah data. Hal ini bertujuan untuk mempermudah dalam pengujian metode k-means dalam mengelompokkan data dengan benar, dan untuk mengetahui keefektifan model k-means segmen 5. Hasil pengujian akan direpresentasikan dalam nilai Purity measure yang mengukur nilai kemurnian dari setiap cluster. Berikut merupakan hasil dari pengujian 2 yang dilakukan pada penelitian ini dapat dilihat pada Tabel 2 dan 3.

*Tabel 2 Hasil Pengujian 2 Segmentasi 5*

Percobaan	Data	K	False	Purity
1	28	4	0	1
2	49	7	2	0.95918
3	70	10	7	0.88571
4	91	13	4	0.95604

*Tabel 3 Hasil Pengujian 2 Nonsegmentasi*

Percobaan	Data	K	False	Purity
1	28	4	2	0.92857
2	49	7	3	0.93877
3	70	10	6	0.9
4	91	13	6	0.93406

Tabel 2 menjelaskan hasil pengujian yang telah dilakukan pada citra dengan segmentasi 5, dan Tabel 3 menjelaskan hasil pengujian yang telah dilakukan pada citra Nonsegmentasi. Hasil dari pengujian ini membuktikan bahwa hasil pengelompokan terhadap citra dengan segmentasi 5 menghasilkan nilai yang lebih bagus daripada citra tanpa segmentasi. Dimana dalam 4 kali percobaan citra dengan segmentasi 5 berhasil mengelompokkan citra lebih bagus daripada citra dengan tanpa segmentasi. Sehingga pada skenario 2 dapat disimpulkan hasil clustering pada model segmen 5 lebih baik daripada citra tanpa segmentasi.

### 3.3 Pengujian 3

Pada pengujian 3 dilakukan untuk mengetahui pengaruh ekstraksi ruang warna RGB dan HSV terhadap penelitian ini. Penelitian dilakukan dengan membandingkan nilai purity pada tabel 2 dan 3 berikut hasil dari pengujian 3 yang disajikan pada Tabel 4.

*Tabel 4 Hasil Pengujian 3 Perbandingan*

Pengujian	RGB		HSV	
	False	Purity	False	Purity
1	0	1	0	1
2	2	0.95918	2	0.95918
3	7	0.88571	7	0.88571
4	18	0.80219	4	0.95604

Tabel 4 menjelaskan bahwa ruang warna RGB dan HSV memiliki pengaruh yang tidak terlalu signifikan dimana nilai HSV memiliki hasil yang lebih bagus daripada ruang warna RGB.

### 4. Kesimpulan

Berdasarkan pengujian yang telah dilakukan, model Clustering K-means dengan segmentasi terbukti mampu mengelompokkan citra liburan dengan baik dari pada model clustering K-means tanpa dilakukannya segmentasi. Hal ini dapat dibuktikan dengan melakukan 4 kali pengujian dengan jumlah data yang berbeda disetiap pengujiannya, nilai purity yang dihasilkan disetiap pengujian cukup signifikan dimana 3 dari 4 pengujian menunjukkan model clustering K-mean dengan segmentasi memiliki nilai purity lebih tinggi dari pada tanpa segmentasi. Hal tersebut terbukti bahwa model clustering K-means dengan segmentasi ini cukup baik dalam mengelompokkan citra. Ekstraksi ruang warna ikut diuji untuk mengetahui pengaruh dalam penelitian ini. Berdasarkan pengujian yang telah dilakukan terbukti bahwa bentuk ekstraksi ruang warna mempengaruhi dalam penelitian ini. Hal ini dapat dibuktikan dengan membandingkan hasil nilai purity dari ruangwarna RGB dan HSV. Hasil uji menunjukkan ruang warna HSV lebih baik daripada ruang warna RGB namun tidak terlalu signifikan. Hal ini dapat disimpulkan bahwa model yang diusulkan cukup baik dalam mengelompokkan citra liburan dari sistem klasterisasi citra.

### Referensi

- [1] H. A. Devy, "Pengembangan Obyek Dan Daya Tarik Wisata Alam Sebagai Daerah Tujuan Wisata Di Kabupaten Karanganyar," *J. Sociol. DILEMA*, vol. 32, no. 1, pp. 34–44, 2017.
- [2] I. P. G. Parma, "Kontribusi Pariwisata Alternatif Dalam Kaitannya Dengan Kearifan Lokal dan Keberlangsungan Lingkungan Alam," *J. Media Komun. FIS Univ. Pendidik. Ganesha*, vol. 9, no. 2, pp. 45–57, 2010.
- [3] N. Asma, L. P. Utomo, and I. A. Saputra, "Partisipasi masyarakat dalam pengelolaan obyek wisata pantai nalera di desa uwevolo kecamatan sinu kabupaten parigi moutong," *E-Journal Geo-Tadulako UNTAD Email.*, 2016.
- [4] A. Trisnawan and W. Harianto, "Klasifikasi Beras Menggunakan Metode K-Means Clustering Berbasis Pengolahan Citra Digital," vol. 1, no. 1, pp. 16–24.
- [5] M. N. Reza, I. S. Na, S. W. Baek, and K. H. Lee, "Rice Yield Estimation Based on K-means Clustering with Graph-cut Segmentation Using Low-altitude UAV Images," *Biosyst. Eng.*, vol. 177, no. 2018, pp. 109–121, 2019.
- [6] K. Tian, J. Li, J. Zeng, A. Evans, and L. Zhang, "Segmentation of Tomato Leaf Images Based on Adaptive Clustering Number of K-means Algorithm," *Comput. Electron. Agric.*, vol. 165, no. March, p. 104962, 2019.
- [7] S. Agustina, D. Yhudo, H. Santoso, N. Marnasusanto, A. Tirtana, and F. Khusnu, "Clustering Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode K-Means Algoritma," *Clust. K-Means*, 2012.
- [8] M. Anggara, H. Sujiani, and N. Helfi, "Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokkan Member Di Alvaro Fitness," *J. Sist. dan Teknol. Inf.*,

- 
- vol. 1, no. 1, pp. 1–6, 2016.
- [9] R. C. Wihandika and M. A. Fauzi, “Implementasi Algoritma Improved K-Means Pada Portal Jurnal Implementasi Algoritma Improved K-Means Pada Portal,” no. April 2018, pp. 1–9, 2017.
- [10] R. Handoyo, R. Rumani, and S. M. Nasution, “Perbandingan Metode Clustering Menggunakan Metode Single Linkage Dan K-Means Pada Pengelompokan Dokumen,” *JSM STMIK Mikroskil*, vol. 15, no. 2, pp. 73–82, 2014.
- [11] “Applications,” p. 167.
- [12] J. Lv, H. Ni, Q. Wang, B. Yang, and L. Xu, “A Segmentation Method of Red Apple Image,” *Sci. Hortic. (Amsterdam)*, vol. 256, no. July, p. 108615, 2019.