

PENERAPAN ALGORITMA K.MEANS UNTUK PENENTUAN PENCOCOKAN PEWARNAAN CLUSTERING SECARA OTOMATIS PADA PRODUK FASHION

Instianti Elyana

Program Studi Manajemen Informatika
Akademik Manajemen Informatika dan Komputer Bina Sarana Informatika
AMIK BSI BOGOR
Jl Merdeka 168 Bogor 16114
instianti.iny@bsi.ac.id

ABSTRACT

Many the needs of the people especially to find products fashion being suitable and appropriate using a web to the needs of user but many web market products fashion make levels color is needed to looking for a match with the skin and the proportion of the body .The need for the measurement of levels color for it was very needed to search products fashion required with classifications uses the method algorithm k-means to get accuracy color being suitable and to clustering staining approaching search products fashion for example by color selection to model suit that in conformity with the wish .It is expected that with this research will help the people to determine match use of color in the fashion desired .The research results show that this method can reduce ketidakakuratan 25 % , so that color later is produced and same in image color will look homogeneous and this study shows the percentage Failure or error rate by more or less under 20 %.

Keyword : Algoritma K.Means, Coloring and fashion Product

I. PENDAHULUAN

Pertumbuhan teknologi web pada saat ini banyak digunakan dan dibutuhkan dalam kehidupan sehari-sehari. Banyaknya perusahaan maupun perorangan yang memasarkan produknya melalui web. Layanan belanja yang semakin banyak mengakibatkan beralihnya layanan belanja yang biasa manual beralih menggunakan web walaupun tidak semuanya beralih menggunakan web. Dalam dunia fashion menempati porsi terbesar dalam layanan belanja berbasis web yang membuat web khusus fashion menyediakan warna fitur browsing yang khas dan utama dalam tampilannya menggunakan visual dan berperan penting dalam memutuskan untuk pembelian.

Gambar produk fashion penting untuk meningkatkan kualitas pencarian. Dari segi warna pun sangat disesuaikan dalam pencarian di web browsing. Pada gambar diketahui warna 24 bit dan diketahui pula jumlah warna yang unik biasanya akan melebihi setengah dari ukuran gambar dan mencapai 16 juta warna. Manusia hanya dapat mengenali 30 warna dan hanya sebagian warna saja. Segmentasi gambar warna berwarna bertujuan untuk menganalisa gambar dan pengenalan pola. Segmentasi gambar berwarna merupakan proses

mempartisi sebuah gambar menjadi beberapa daerah yang homogen atas dasar gambar persamaan karakteristik tertentu.

Seseorang yang berpenampilan baik selain harus memiliki tingkah laku dan kepribadian yang baik juga harus didukung oleh penampilan luar berupa keindahan dan ketepatan dalam berbusana. Ketepatan busana yang baik sebaiknya memiliki ketepatan dari segi pola dan warna yang sesuai serta desain yang sesuai bagi pengguna. Dari penjelasan latar belakang maka dapat dirumuskan beberapa permasalahan, sebagai berikut:

1. Adanya kesulitan pengguna dalam memadupadankan busana dengan warna sebagai kebutuhan penggunaan busana. Kesulitan tersebut berupa penggunaan busana yang disesuaikan dengan usia, warna kulit, kesempatan, musim dan juga ukuran badan.
2. kebanyakan orang kesulitan dalam mencari busana yang memiliki ukuran warna sesuai dengan bentuk tubuh dan kesesuaian warna dengan kulit.

Tujuan dari penelitian ini adalah untuk merancang dan merealisasikan ALgoritma K-Means analisis untuk menentukan kualitas warna. Desain pemilihan warna busana dirancang berdasarkan variabel yang

mempengaruhi solusi penggunaan busana yang berupa solusi bentuk model busana dengan mencocokkan warna yang sesuai dengan jenis kulit serta penggunaan bahan kain yang disesuaikan dengan musim dan kesempatan penggunaan berdasarkan variabel yang diinputkan oleh pengguna serta dapat memberikan perhitungan perkiraan bahan dan warna yang akan digunakan.

II. KAJIAN LITERATUR

Data clustering merupakan salah satu metode Data Mining yang bersifat tanpa arah. Ada dua jenis data clustering yang sering dipergunakan dalam proses pengelompokan data yaitu hierarchical (*hirarki*) data clustering dan non-hierarchical (*no hirarki*) data clustering. K-Means merupakan salah satu atau lebih clustering non hirarki yang berusaha mempartisi data kedalam *cluster* atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster atau kelompok sehingga data yang memiliki karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain.

Adapun tujuan dari data *clustering* ini adalah untuk meminimalisasikan *object function* yang akan diset dalam proses clustering pada umumnya meminimalisasikan variasi didalam suatu cluster dan memaksimalkan variasi antar cluster.

A. Metode K-Means Clustering

K-Means merupakan metode *non heirarchical* yang pada awalnya mengambil sebagian dari banyaknya komponen dari populasi untuk dijadikan pusat cluster dipilih secara acak dari sekumpulan data warna. Selanjutnya K-Means menguji masing-masing komponen didalam populasi data dan menandai komponen tersebut kesalah satu pusat cluster yang telah didefinisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap pusat cluster.

B. Algoritma K-means Analisis

K-Means adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervise dan merupakan salah satu metode yang melakukan pengelompokan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain.

Pada dasarnya dapat diterapkan pada permasalahan dalam memahami perilaku konsumen, mengidentifikasi peluang produk baru dipasaran dan Algoritma K-Means ini juga dapat digunakan untuk meringkas objek dari jumlah besar sehingga lebih memudahkan untuk mendeskripsikan sifat-sifat atau karakteristik dari masing-masing kelompok.

Secara umum metode *K-Means Cluster Analisis* menggunakan Algoritma Sebagai berikut:

1. Tentukan K sebagai jumlah cluster yang di bentuk. Untuk menentukan banyaknya cluster k dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan berapa banyak cluster.

2. Tentukan k *Centroid* (titik pusat cluster) awal secara random. Penentuan centroid (titik pusat cluster) awal secara random. Penentuan centroid awal dilakukan secara random atau acak dari objek-objek yang tersedia sebanyak K Cluster, kemudian untuk menghitung centroid cluster ke – I berikutnya

C. Data mining

Data mining merupakan kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santoso, 2007). Tujuan utama dari klastering adalah pengelompokan sejumlah data atau obyek ke dalam klaster akan berisi data yang semirip mungkin (Santosa, 2007).

III. METODE PENELITIAN

A. Algoritma Fuzzy K-Means

Algoritma FCM adalah sama dengan teknik hillclimbing digunakan untuk teknik clustering untuk segmentasi gambar. Algoritma FCM bertujuan membagi setiap pixel menjadi koleksi dari M Fuzzy cluster centroid dengan memberikan beberapa kriteria N adalah jumlah pixel pada gambar dan M adalah eksponensial derajat keanggotaan dalam persamaan:

$$W_m(U, C) = \sum_{i=1}^N \sum_{j=1}^M u_{ji}^m d_{ji}^2,$$

Dimana U_{ji} adalah derajat keanggotaan i pixel ke j cluster centroid, $U_i = (U_{i1}, U_{i2}, \dots, U_{im})$ adalah derajat keanggotaan i pixel diasosiasi dengan setiap cluster centroid. $U = (U_1, U_2, \dots, U_N)$ adalah matrik derajat keanggotaan dan $C = (c_1, c_2, \dots, c_m)$ adalah cluster centroid.

Derajat kekompakan dan keseragaman cluster centroid sangat tergantung pada fungsi objektif FCM Umumnya semakin kecil fungsi FCM mengindikasikan kekompakan dan keseragaman cluster centroid. FCM digunakan untuk meningkatkan kekompakan pada cluster Algoritmanya sebagai berikut:

3. Memilih iterasi akhir thresholding. ϵ adalah jumlah positif terkecil pada range $[0,1]$ dan jumlah iterasi q ke 0.
4. Menghitung $U(q)$ menurut $C(q)$ dengan formula Persamaan.

$$u_{ji} = \frac{1}{\sum_{k=1}^M \left(\frac{d_{ji}}{d_{ki}}\right)^{2/(m-1)}},$$

dimana $1 \leq j \leq M$ dan $1 \leq i \leq N$, jika $d_{ji} = 0$ kemudian $u_{ji} = 1$ dan pilih derajat keanggotaan lain pada pixel ke 0.

5. Langkah ketiga, hitung $C_{(q+1)}$ berdasarkan $U_{(q)}$ pada Persamaan.

$$c_j = \frac{\sum_{i=1}^N u_{ji}^m x_i}{\sum_{i=1}^N u_{ji}^m},$$

dimana $1 \leq j \leq m$. Langkah keempat, perbaharui $U_{(q+1)}$ berdasarkan $C_{(q+1)}$ berdasarkan persamaan (8). Kemudian bandingkan $U_{(q+1)}$ dengan $U_{(q)}$, jika $U_{q+1} - U_q \leq \epsilon$ maka berhenti iterasi. Lainnya jika $q = q + 1$ dan ulangi langkah 2

sampai langkah 4 sampai $U_{q+1} - U_q > \epsilon$.

- B. Penelitian ini dilakukan dengan langkah-langkah sebagai berikut :
 1. Desain palet dimana fase digunakan untuk memilih sekumpulan warna yang dianggap mampu mewakili warna asli. Setiap warna terpilih mengandung 3 dimensi untuk warna RGB yang bisa dianalogikan sebagai kode dalam sebuah buku kode, dimana palet adalah buku kode.
 2. Pemetaan piksel dimana fase untuk menandai setiap piksel sesuai desain palet. Setiap piksel pada citra asli dipetakan ke warna terdekat dalam palet dengan cara menemukan warna yang sesuai atau yang paling dekat dari palet.

Tujuan dari fase ini adalah untuk mengurangi jumlah warna unik N' dari citra menjadi K sehingga $N' \ll K$ dengan distorsi sekecil mungkin. Pada banyak aplikasi, citra asli dengan piksel 24 bita akan dikurangi menjadi 8 bita atau lebih kecil.

Warna menjadi 32 warna (nilai $k = 32$) dengan menggunakan skema inisialisasi untuk semua warna. Kondisi terminasi dilakukan setelah 20 kali iterasi dengan ambang konvergensi dan keefektifitas metode kuantisasi dinyatakan dengan menggunakan pengukuran nilai mean square error (MSE) :

$$MSE(X, \tilde{X}) = \frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W \|x(h, w) - \tilde{x}(h, w)\|_2^2$$

MSE menyatakan rata-rata distorsi dan umum digunakan untuk mengevaluasi warna .

IV. PEMBAHASAN

A. Data warna dites

Kumpulan data warna yang digunakan dites yang biasa dipakai dalam literature ini yaitu warna merah (29% warna utama), biru (58 % warna utama), kuning (31 %), hijau (57 % warna utama), coklat (13 %), ungu (42% warna utama), abu-abu (47 % warna utama), dan hitam (7 warna unik).

Dalam penelitian ini dilakukan segmentasi dengan tahap :

1. Membagi warna menjadi 3 bagian dicari histogram dengan modifikasi pada 3 warna yaitu merah, hijau dan biru. Pada masing-masing gambar kemudian dilakukan

- pengurangan peak dan valley yang nilainya jauh dari nilai thresholdnya.
- 2. Inialisasi Regional dalam 3 warna
- 3. Penggabungan 3 warna atau merging berupa cluster

K-means merupakan salah satu metode data clustering non hirarki yang digunakan untuk mengklasifikasi data yang ada ke dalam bentuk satu atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok atau cluster, sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain.

Data clustering menggunakan metode k-means ini secara umum dilakukan dengan algoritma dasar sebagai berikut :

1. Tentukan jumlah cluster
2. Alokasikan data ke dalam cluster secara random
3. Hitung rata-rata dari data yang ada di masing-masing cluster
4. Alokasijian masing-masing data ke rata-rata terdekat
5. Kembalikan ke step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai rata-rata ada yang diatas nilai threshold yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan di atas nilai threshold yang ditentukan.

B. Mendefinisikan Indeks, parameter dan variable dalam penelitian.

Pada bagian berikut ini akan didefinisikan indeks, parameter dan variable dari metode yang akan digunakan dalam penelitian ini :

Indeks :

1. i data ke- ($i= 1, 2, 3, \dots, m$)
2. j variable ke- ($j= 1, 2, 3, \dots, n$)
3. k jumlah cluster
4. m jumlah data
5. n jumlah variabel
6. r cluster ke- ($r = 1, 2, 3, \dots, k$)

Parameter :

1. W_j bobot variable ke- ($j = 1, 2, 3, \dots, n$)
2. α parameter situasi ($0 \leq \alpha \leq 1$)
3. a_i nilai agregat data ke- ($i= 1, 2, 3, \dots, m$)
4. C_r pusat cluster ke- dengan $r = 1, 2, \dots, k$

Variabel :

1. X_{ij} Data ke= i pada variable ke- j ($i=1, 2, 3, \dots, m$; dan $j = 1, 2, 3, \dots, n$)
2. μ_j mean dari variable ke- ($j=1, 2, 3, \dots, n$)
3. μ_r mean dari klaster ke- ($r= 1, 2, \dots, k$)

Suatu metode pengujian konektivitas piksel dengan tetangganya dan pemberian nama dari piksel-piksel yang telah dihubungkan. Salah satunya algoritma yang digunakan dalam pemberian nama piksel

C. Teknik pemetaan warna

Teknik pemetaan warna merupakan bagian salah satu teknik segmentasi pewarnaan merupakan bagian salah satu teknik segmentasi pewarnaan menggunakan metode klasterisasi. Hal ini disebabkan dalam pemetaan warna masukkam akan dikelompokkan sesuai dengan kesama-kesamaan warna yang dimiliki. Sehingga tahap-tahap yang akan digunakan mempunyai kesamaan dengan metode klasterisasi.

Cluster adalah kesatuan nilai dalam jarak tertentu pada kepadatan suatu daerah relative besar dibandingkan dengan kepadatan nilai daerah sekitarnya. Teknik clustering bermanfaat untuk segmentasi citra dan klarifikasi data yang belum diolah untuk menciptakan kelas-kelas.

Warna diwakili dalam vector 3 dimensi dari nilai titiknya. Masing-masing komponen biru (GRB). Jika dua vektor saling berdekatan, warna akan ditampilkan serupa, rata-rata dari dua garis vector, jika warna yang akan ditampilkan sangat berbeda. Dan akan diambil jalan tengah dengan menghadirkan suatu segmen warna secara kasar dari warna aslinya. Adapun cara-cara lain yang tidak membatasi untuk menghadirkan suatu warna dengan garis vector 3 dimensi.

Penampilan warna mempengaruhi hasil proses klasterisasi:

1. Menetapkan data set dari algoritma yang akan digunakan K-Means dengan melakukan pengambilan nilai acak dari k
2. Penampilan RGB dari tiap pixel diciptakan dan menghasilkan dataset dalam 3 vektor
3. Algoritma K-Means diterapkan pada dataset, menetapkan klasterisasi pusat k Algoritma K-Means akan menghadirkan k warna untuk menggambarkan pewarnaan

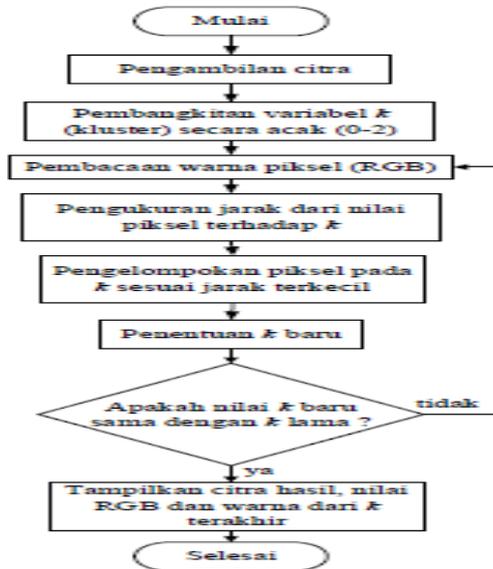
4. Tiap-tiap piksel citra dikonversi dalam suatu garis vector RGB dan ditampilkan

5. Menggunakan rata-rata dari kelompok warna yang dihasilkan

D. Pembagian Variabel K (cluster) secara acak.

yang akan diinginkan. Penentuan nilai parameter k (nilai RGB nya) akan dilakukan secara acak

Warna atau citra yang akan dilakukan segmentasi, pertama-tama ditentukan jumlah k



Sumber : Hasil Penelitian (2015)

Gambar 1. Penentuan Nilai Parameter

E. Pengelompokkan Piksel pada k sesuai jarak terkecil

$$\frac{\text{Jumlah data yang sesuai label aslinya}}{\text{Jumlah data keseluruhan}} \times 100\%$$

Pengelompokkan piksel ini dilakukan agar nantinya dalam pemetaan warna yang dimiliki jarak terkecil atau kemiripan warna terbesar dapat benar-benar menjadi satu kelompok. Proses ini merupakan proses awal dari segmentasi citra yang bertujuan salah satunya untuk memperjelas batas objek sesuai dengan k yang diinginkan.

Memasukkan nilai k yang akan digunakan dengan jumlah cluster warna 24, 64, 128 dan 256.

F. Pengukuran Accuration Rate

Nilai accuration rate merupakan perbandingan antara jumlah cluster yang sesuai dengan label pada data aslinya dibandingkan dengan jumlah data keseluruhan.

Accuratioan rate =

Jumlah data keseluruhan

G. Teknik analisis data

Dalam tahapan penelitian ini dilakukan menggunakan bantuan matlab 7.01, dengan cara membuat program.

Sederhana (*matlab code*) yang dapat digunakan secara tepat untuk memperoleh hasil clustering pada masing-masing metode yang digunakan. Dan program tersebut kemudian diaplikasikan pada matlab command window dengan memasukkan input: data yang dianalisis (x), jumlah cluster yang diinginkan (k), dan jumlah iterasi yang digunakan (i). Sehingga proses analisis data pada penelitian ini diharapkan lebih efisien waktu dan hasilnya lebih akurat.

Tabel 1. Peringkat nilai pengukuran warna dengan beberapa inisialisai

k	Merah	biru	kuning	hijau	Mean
32	189.7	113.6	66.95	40.57	102.7
64	173.5	104.7	62.98	39.38	95.14
128	160	93.7	57.43	35.87	86.72
256	157.7	91.28	55.16	43.26	84.6%

Sumber : Hasil Penelitian (2015)

Tabel 2. Peringkat waktu komputasi Cpu dan Algoritma

k	coklat	unggu	abu-abu	hitam	Mean
32	1.46	1.93	3.01	4.99	2.84
64	1.5	2.38	3.82	7.05	3.68
128	1.31	1.9	2.67	4.78	2.66
256	2.93	5.13	9.46	17.7	8.8 %

Sumber : Hasil Penelitian (2015)

Tabel 3. Tabel hasil kesesuaian

k	Merah	biru	kuning	hijau	Mean
32	2.1	3.2	22.1	31.1	21.5
64	1.4	4.40	3.00	1.11	19.2
128	1.31	3.11	4.11	2.11	2.00
256	1.3	2.90	3.33	18.3	20.00 %

Sumber : Hasil Penelitian (2015)

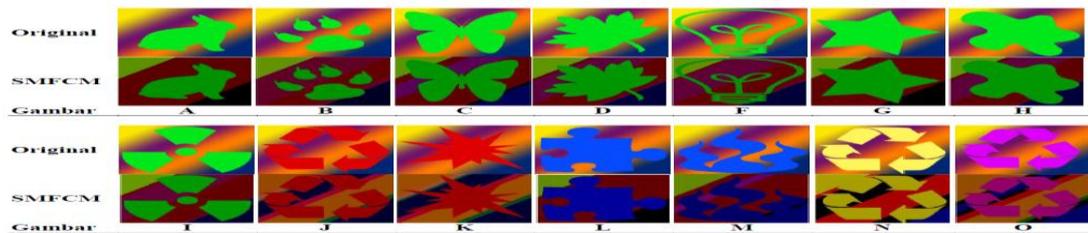
Tabel 3. Tabel hasil kesesuaian

k	Merah	biru	kuning	hijau	Mean
32	1.2	3.1	4.11	3.11	3.11
64	2.1	3.2	5.11	7.68	3.2
128	4.13	2.00	8.12	7.75	5.11
256	6.21	6.13	7.23	9.00	25.00 %

Sumber : Hasil Penelitian (2015)

Error rate akan didapatkan dari pemisahan warna antara warna utama dan background warna yang dipilih. Hal ini dikarenakan memiliki derajat

warna yang sama contohnya derajat warna kemerahan dan warna kebiruan bisa dilihat digambar dibawah ini :



Sumber : Hasil Penelitian (2015)

Gambar 2. Hasil segmentasi pada gambar sintesis

Dari tabel 1 dan 2 dilihat dari perspektif kualitas visual warna terkuantisasi, skema inialisasi dengan K-means secara konsisten menghasilkan kualitas yang lebih baik jika dibandingkan dengan yang lain dan waktu komputasi menghasilkan cpu yang tidak konsisten. Hal ini disebabkan karena karena setiap kali iterasi pencarian warna akan melakukan pemilihan warna utama walaupun pemilihan warna diambil secara acak.

V. PENUTUP

Penelitian tentang pengukuran pewarnaan ini menggunakan teknik pemetaan warna dan menghasilkan:

1. Semakin besar ukuran citra warna maka waktu iterasi yang dibutuhkan semakin lama dan hal ini disebabkan oleh pengelompokan antara piksel-pikselya yang banyak. Namun ukuran citra pewarnaan asli tidak banyak berpengaruh terhadap jumlah iterasi.
2. Semakin banyak cluster maka jumlah dan waktu iterasi juga akan semakin banyak atau lama. Hal ini disebabkan oleh proses perhitungan dan percobaan diantara clusternya saat iterasi juga semakin banyak dan lama.
3. Waktu selama proses iterasi dipengaruhi oleh histogram
4. Waktu selama proses dipengaruhi oleh sebaran histogram. Dari penelitian yang dilakukan waktu selama proses iterasi yang dibutuhkan cenderung lebih lama pada pewarnaan

DAFTAR PUSTAKA

- Agusta Y. (2007). *K-means, penerapan, permasalahan dan metode Terkait, jurnal system dan informatika* Vol,3
- Chintalapalli, M., *Color Image Segmentation by Clustering*, <http://www.cis.temple.edu/..../Courses/CIS601-04/Projects/Manjit/Color%20Image%20segmentation%20by%20Clusteri%20doc%20april%202009>
- Munir, Rinaldi. (2009). *Pengolahan Citra Digital Dengan Pendekatan Algoritmik*, Informatika, Bandung,
- Prahasta, E. (2005). *Sistem Informasi Geografis, Informatika*, Bandung,
- Y.-C. Hu, M.-G. Lee, *K-means based color palette design scheme with the use of stable flags*, *Journal of Electronic Imaging* 16 (3) (2007) 033003.