

Seleksi Fitur *Information Gain* Pada Temu Kembali Citra Jenis Makanan Menggunakan *Dominant Color Descriptor* Dan *Gray Level Co-occurrence Matrix*

Sulaiman Triarjo¹, Yuita Arum Sari², Sigit Adinugroho³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹s.triar@live.com, ²yuita@ub.ac.id, ³sigit.adinu@ub.ac.id

Abstrak

Informasi nutrisi di dalam media sosial ditunjang oleh citra makanan yang sedang diulas. Diperlukan usaha keras bila ingin menelusuri makanan-makanan yang sejenis yang mempunyai nutrisi yang hampir sama. Oleh karena itu dibutuhkan sistem pencarian informasi untuk mempercepat proses pencarian informasi. Penelitian ini menggunakan metode *Dominant Color Descriptor* (DCD) untuk ekstraksi fitur warna dan metode *Gray Level Co-occurrence Matrix* (GLCM) untuk ekstraksi fitur tekstur serta seleksi fitur *information gain* untuk menyeleksi fitur tekstur sehingga tidak semua fitur tekstur digunakan. Data yang digunakan berupa 29 jenis citra makanan dengan total 435 citra dengan setiap jenis makanan 15 citra. Pengujian dilakukan dengan membandingkan performa perhitungan jarak *Euclidean distance*, *Chebyshev distance*, dan *Manhattan distance* untuk fitur tekstur dan *Quadratic distance* dan *Yang distance* untuk fitur warna. penggunaan evaluasi nilai MAP, hasil pengujian menggunakan fitur tekstur didapatkan nilai MAP sebesar 0,5542 dengan jarak *Euclidean* dan tanpa adanya seleksi fitur. Hasil pengujian menggunakan fitur warna didapatkan nilai MAP sebesar 0,7488 pada saat menggunakan jarak *Yang*. Pengujian menggunakan fitur tekstur warna dan fitur tekstur didapatkan nilai sebesar 0,7118 dengan menggunakan jarak *Manhattan* dan *Yang* dengan 10 fitur. Pada penelitian ini penggunaan DCD lebih efektif dari pada GLCM dengan menghasilkan nilai MAP yang lebih tinggi.

Kata kunci: temu kembali informasi, makanan, warna, tekstur, *Dominant Color Descriptor*, *Gray Level Co-occurrence Matrix*, *information gain*

Abstract

Nutritional information on social media is supported by image of food being reviewed. It requires hard work to explore similar foods that have almost same nutrition. Therefore an information search system is needed to speed up the information search process. This research has been conducted to be able to search for similar informations based on a query in form of image. It uses Dominant Color Descriptor method for color feature extraction and Gray Level Co-occurrence Matrix method for texture feature extraction and information gain selection feature to select texture features. The data used were 29 types of food imaged with total is 435 images which each type has 15 images. Testing is done by comparing the performance of calculation of Euclidean distance, Chebyshev distance, and Manhattan distance for texture feature and Quadratic distance and Yang distance for color feature. The evaluation uses MAP value, test result using only the texture feature obtained MAP value of 0,5542 using Euclidean distance and without feature selection. The test result using only color feature obtained MAP value of 0,7488 when using Yang distance. And testing using color feature and texture feature obtained a value of 0,7118 by using Manhattan distance and Yang distance with 10 features. In this research, the use of DCD was more effective than GLCM by producing higher MAP value.

Keywords: content-based information retrieval, food, color, texture, *Dominant Color Descriptor*, *Gray Level Co-occurrence Matrix*, *information gain*

1. PENDAHULUAN

Sebagai akibat dari pertumbuhan dunia

internet yang begitu pesat, peningkatan data dan peningkatan arus informasi juga semakin besar. Peningkatan data dan arus informasi salah

satunya disebabkan oleh penggunaan media sosial. Media sosial adalah sebuah metode komunikasi yang digunakan untuk memfasilitasi interaksi sosial dengan dasar internet (Tobey dan Manore, 2014). Dengan adanya interaksi sosial berarti adanya pertukaran informasi yang dilakukan oleh pengguna media sosial. Informasi yang beredar pada media sosial salah satunya adalah informasi mengenai nutrisi dan gizi makanan. Oleh sebab itu media sosial dapat menjadi pendidik mengenai nutrisi yang cepat dan murah (Tobey and Manore, 2014).

Informasi mengenai nutrisi dan gizi di dalam media sosial tidak terlepas dari sebuah citra makanan yang sedang diulas nutrisinya. Penggunaan citra dimaksudkan agar pembaca dapat mengetahui jenis makanan yang seperti apa yang sedang diulas. Karena informasi di media sosial begitu banyak akan ada beberapa makanan yang berbeda namun memiliki jenis yang sama atau memiliki informasi nutrisi yang sama. Seseorang akan membutuhkan usaha keras untuk mencari makanan yang mempunyai jenis makanan yang sama di dalam sebuah media sosial terlebih lagi pengguna tersebut tidak tahu nama makanan yang ingin dicarinya. Oleh sebab itu diperlukan sistem pencarian untuk memperoleh informasi dari data yang sangat besar tersebut. Salah satu sistem yang berguna untuk mendapatkan informasi adalah sistem temu kembali informasi berdasarkan citra. Sistem temu kembali informasi berdasarkan citra sangat luas jangkauannya mulai citra satelit, citra kendaraan, citra wajah seseorang, citra sidik jari, citra mainan, hingga citra makanan. Sistem temu kembali informasi berdasarkan citra merupakan pekerjaan dalam memperoleh citra-citra yang relevan dari kueri yang diberikan pengguna yang berupa citra. Penggunaan kueri yang berupa citra lebih berguna daripada menggunakan kueri yang berupa kata maupun kalimat (Rashno & Sadri, 2017). Pada penggunaan kueri yang berupa kata atau kalimat sangat dipengaruhi oleh metadata yang ada pada citra di dalam database. Akan membutuhkan kerja yang sangat berat untuk seseorang memberikan tag yang men-detail terkait jenis dan keterangan objek yang ada pada sebuah citra. Sebaliknya, dengan menggunakan kueri yang berupa citra, dapat diekstrak fitur-fitur yang ada pada citra yang kemudian dicari kemiripannya dengan citra-citra yang ada di dalam basis data.

Penelitian-penelitian yang berkaitan dengan temu kembali informasi berdasarkan citra telah banyak dilakukan seperti yang dilakukan Li dan

Bao, (2010) yang membandingkan metode yang digunakan untuk sistem temu kembali informasi berdasarkan citra. Dengan usulan metode yang menggunakan *Dominant Color Descriptors* dan metode yang dibandingkan menggunakan *Merged Histogram*. Dengan kriteria evaluasi yang digunakan menggunakan *Average Normalized Modified Retrieval Rank* (ANMRR) dan *Average Retrieval Rate* (ARR). Hasil yang diperoleh menunjukkan bahwa metode usulan yang digunakan memperoleh hasil yang lebih baik dari metode yang dibandingkan.

Penelitian yang dilakukan oleh Rasli et al., (2012) membuat sistem temu kembali informasi citra dengan menggunakan fitur tekstur, lokasi, bentuk, dan warna. Fitur-fitur yang digunakan didapatkan melalui penggunaan metode *color histogram* dengan menggunakan *gray level co-occurrence matrix* dan metode *color histogram* dengan menggunakan *K-Means*. Dengan menggunakan perhitungan jarak *euclidean distance* dihitung kemiripan antara kueri dengan datanya sudah di latih. Dari hasil yang didapatkan, metode *color histogram* dengan menggunakan *gray level co-occurrence matrix* mendapatkan nilai yang lebih tinggi pada fitur tekstur dan lokasi.

Penelitian yang dilakukan oleh Aini, Sari, & Arwan, (2018) meneliti efektifitas dari algoritme seleksi fitur *Information Gain* dengan proses klasifikasi menggunakan metode *KNN* untuk atribut numerik dan *Naïve Bayes* untuk atribut kategori. Penggunaan seleksi fitur *information gain* digunakan karena masih terdapat fitur-fitur yang tidak relevan dan untuk mengurangi dimensi atribut sehingga proses komputasi menjadi lebih cepat dan hasil yang didapatkan lebih akurat. Hasil dari penelitian ini menunjukkan bahwa pemakaian seleksi fitur *information gain* pada proses klasifikasi tersebut memberikan hasil yang lebih akurat dari pada tidak menggunakan seleksi fitur *information gain*

Berdasarkan berbagai penelitian yang mengangkat topik sistem temu kembali informasi tersebut. Penelitian sistem temu informasi citra jenis makanan dengan ekstraksi fitur warna menggunakan *Dominant Color Descriptor* (DCD) dan fitur tekstur menggunakan *Gray Level Co-occurrence Matrix* (GLCM) kemudian akan dilakukan seleksi fitur *information gain* untuk menyeleksi fitur-fitur yang didapatkan.

2. METODE PENELITIAN

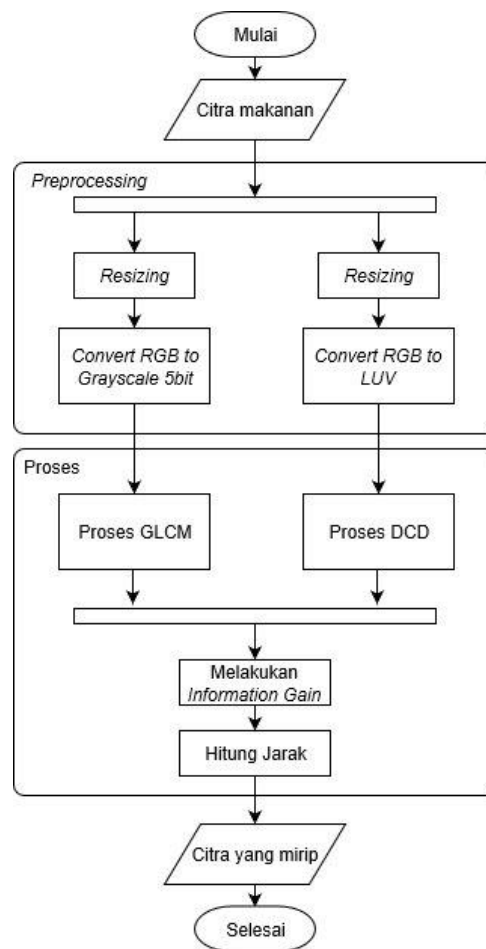
2.1 Data Yang Digunakan

Data yang digunakan berupa citra makanan yang didapatkan dengan cara memfoto makanan dengan menggunakan *smartphone*. Banyak jenis makanan adalah 29 jenis, yaitu: donat, roti gandum, roti tawar, indomie goreng, mie gepeng goreng, telur ceplok, telur dadar, *fried chicken*, rendang, mentimun, selada air, kubis, selada, kemangi, tomat, stroberi, pisang hijau, pisang kuning, jeruk oranye, jeruk ijo, nasi kuning, nasi merah, Oreo, Beng-Beng, Soba Mie, Tim Tam, Happy Toss, Gerry Saluut, Biskuat coklat, Milo nuggets, dan Genji Pie. Setiap jenis makanan diambil secara acak sebanyak 9 citra dari dataset hasil pengambilan citra untuk menjadi data latih sehingga total ada 279 data. Sedangkan untuk data uji diambil 3 citra secara acak setiap jenis makanan sehingga total ada 93 data. Pengambilan data dilakukan pada tanggal 28 Agustus 2018 di ruang laboratorium riset komputasi cerdas yaitu di ruang F9.3 gedung F Fakultas Ilmu Komputer Universitas Brawijaya. Ruang laboratorium riset Penelitian dilakukan pada pukul 11.00 WIB dan berakhir pada pukul 14.00 WIB.

2.2 Rancangan Algoritma

Penelitian dimulai dengan melakukan *preprocessing* pada citra data latih. Setelah *preprocessing* dilakukan, selanjutnya dilakukan ekstraksi fitur warna dengan metode DCD. Kemudian dilakukan ekstraksi fitur tekstur dengan menggunakan metode GLCM. Setelah fitur-fitur berhasil didapatkan, selanjutnya fitur-fitur tersebut disimpan pada basis data supaya tidak melakukan ekstraksi fitur data latih kembali. Setelah itu dilakukan perhitungan seleksi fitur *information gain* dan kembali menyimpan atribut-atribut hasil seleksi *infotmation gain*.

Setelah selesai mengekstraksi fitur pada data latih, selanjutnya melakukan pengujian dengan melakukan *preprocessing* pada citra data uji. Setelah itu dilakukan ekstraksi fitur yang sama pada data latih. Selanjutnya dihitung kemiripannya dengan menggunakan perhitungan jarak antara fitur dari kueri dengan fitur yang ada di dalam basis data. Setelah itu diurutkan secara menaik dan dipilih sejumlah data terdekat dengan kueri untuk ditampilkan sebagai hasil. Kemudian hitungan nilai MAP.



Gambar 1 Diagram Alir Penelitian

2.3 Preprocessing

Tahap *preprocessing* bertujuan untuk meningkatkan kualitas citra maupun untuk mempersiapkan citra agar cocok untuk digunakan dengan metode-metode yang akan dilakukan. Di tahap *preprocessing* yang dilakukan antaran lain:

1. *Resizing*

Resizing adalah proses mengubah ukuran piksel sebuah citra baik memperbesar maupun memperkecil. *Resizing* digunakan pada penelitian ini adalah membuat ukuran piksel sebuah citra menjadi ukuran 1024x1024 dengan menggunakan metode *bilinear interpolation*.

2. *Convert Color Space RGB to GrayScale 5 Bit*

Pada proses ini dilakukan sebagai *preprocessing* untuk metode GLCM. Karena proses GLCM membutuhkan 1 nilai yang konkrit, sehingga diperlukan untuk merubah ruang warna dari RGB ke *grayscale*. Menurut OpenCV, (2016) nilai grayscale sebuah citra bisa didapatkan dengan menggunakan persamaan 1.

$$Grayscale_{i,j} = R_{i,j} * 0.299 + G_{i,j} * 0.587 +$$

$$B_{i,j} * 0.114 \tag{1}$$

Hasil awal dari proses merubah adalah citra grayscale dengan 256 tingkat. Sehingga akan ada banyak sekali nilai di dalam matriks GLCM yang bernilai 0. Oleh karena itu diperlukan konversi dari 256 tingkat menjadi 32 tingkat dengan cara membagi setiap nilai piksel citra grayscale.

3. Convert Color Space RGB to CIE LUV

Proses ini merupakan proses preprocessing untuk metode DCD. Pada standar MPEG-7 proses DCD menggunakan ruang warna CIE LUV oleh karena itu diperlukan konversi ruang warna dari RGB menjadi CIE LUV. Proses konversi dilakukan dari merubah citra RGB

$$v'_{i,j} = \frac{9Y_{i,j}}{X_{i,j} + 15Y_{i,j} + 3Z_{i,j}} \tag{5}$$

$$U_{i,j} = 13L(u'_{i,j} - 0.19793943) \tag{6}$$

$$V_{i,j} = 13L(v'_{i,j} - 0.46831096) \tag{7}$$

Keterangan:

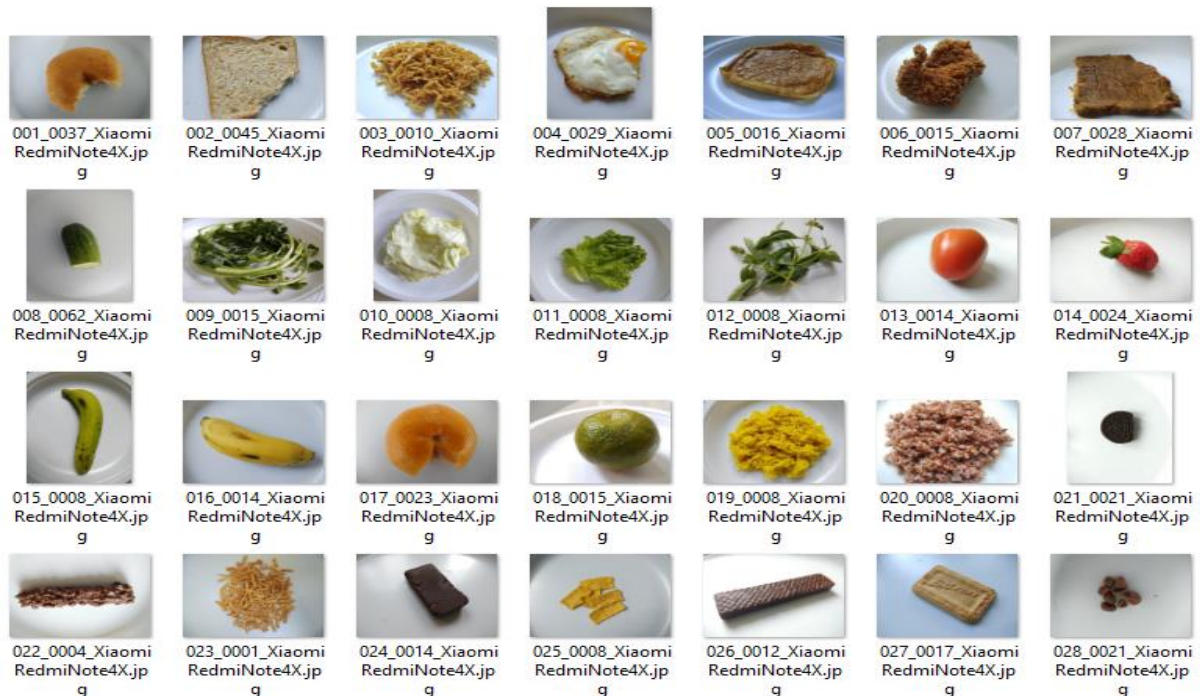
$R_{i,j}$ = channel Red pada citra RGB pada baris i dan kolom j

$G_{i,j}$ = channel Green pada citra RGB pada baris i dan kolom j

$B_{i,j}$ = channel Blue pada citra RGB pada baris i dan kolom j

2.4 Dominant Color Descriptor (DCD)

Pada proses DCD ini sebuah citra akan dikuantisasi nilai-nilai pikselnya menggunakan metode K-Means dengan K = 4. Menurut Fierro-



Gambar 2 Contoh Data Yang Digunakan

menjadi XYZ menggunakan persamaan 2. Kemudian dilanjutkan mencari nilai L untuk setiap piksel dengan persamaan 3. Setelah itu dilanjutkan mencari nilai u' dengan persamaan 4 dan nilai v' dengan persamaan 5. Kemudian mencari nilai U dengan persamaan 6 dan nilai V dengan persamaan 7.

$$\begin{bmatrix} X_{i,j} \\ Y_{i,j} \\ Z_{i,j} \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \cdot \begin{bmatrix} R_{i,j} \\ G_{i,j} \\ B_{i,j} \end{bmatrix} \tag{2}$$

$$L_{i,j} = \begin{cases} 116Y_{i,j}^{\frac{1}{3}} & \text{untuk } Y_{i,j} > 0.008856 \\ 903.3Y_{i,j} & \text{untuk } Y_{i,j} \leq 0.008856 \end{cases} \tag{3}$$

$$u'_{i,j} = \frac{4X_{i,j}}{X_{i,j} + 15Y_{i,j} + 3Z_{i,j}} \tag{4}$$

Radilla, et al., (2013), metode ini menggunakan prosedur-prosedur umum untuk mengekstraksi warna seperti memilih ruang warna, color quantization, histogram calculation, penentuan warna dominan, proses perhitungan persen setiap warna yang dominan. Inisialisasi centroid awal menggunakan algoritme K-Means++, yaitu dengan memilih centroid pertama secara acak kemudian pemilihan centroid berikutnya dilakukan dengan menggunakan persamaan 8 (Bahmani et al., 2012).

$$c_{i+1} = \max\left(\frac{D(x,c_i)^2}{\sum_{x \in X} D(x,c_i)^2}\right) \tag{8}$$

Di mana D(x,ci) adalah jarak sebuah piksel di dalam citra dengan centroid i. Kemudian

dilanjutkan proses K-Means *clustering* seperti yang dilakukan oleh Lloyd.

Setelah dilakukan proses kuantisasi maka akan dibentuk fitur-fitur GLCM seperti pada persamaan 9 (Verdaguer, 2009).

$$F = \{(C_i, P_i)\}, i = 1, 2, \dots, N \quad (9)$$

Di mana N merupakan banyaknya *cluster*. C_i merupakan nilai *centroid* ke i, dan P_i merupakan presentase *cluster* i di dalam citra.

2.5 Gray Level Co-occurrence Matrix (GLCM)

GLCM merupakan sebuah matrix yang menyimpan informasi posisi-posisi piksel-piksel yang mempunyai nilai tingkat keabuan yang mirip (Sharma & Batra, 2014). Matriks GLCM didapatkan dengan cara menghitung semua pasangan piksel-piksel yang dibatasi oleh sebuah vektor perpindahan. Langkah-langkah dalam membuat matriks GLCM adalah sebagai berikut.

1. Menghitung pasangan-pasangan piksel *grayscale* menurut vektor perpindahan yang dilakukan yaitu 0°, 45°, 90°, 135°.
2. Membuat matriks simetris dengan menjumlahkan hasil perhitungan pasangan piksel dengan hasil transposenya.
3. Melakukan normalisasi menggunakan persamaan 10.

$$P_{i,j} = \frac{c_{i,j}}{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} c_{i,j}} \quad (10)$$

Di mana $c_{i,j}$ merupakan nilai sel pada i dan j.

4. Melakukan ekstraksi fitur tekstur. Terdapat 10 fitur tekstur yang diekstraksi pada GLCM ini menurut (Hall-Beyer, 2017), yaitu antara lain:

1. *Contrast*
Contrast digunakan untuk menunjukkan besarnya perbedaan intensitas tingkat keabuan pada selisih indeks kuadrat.

$$Contrast = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{ij} (i - j)^2 \quad (11)$$

2. *Dissimilarity*
Dissimilarity digunakan untuk menunjukkan besarnya perbedaan intensitas tingkat keabuan pada selisih indeks.

$$Dissimilarity = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{ij} (i - j) \quad (12)$$

3. *Homogeneity*
Homogeneity digunakan untuk mengukur nilai yang naik dengan *contrast* yang sedikit.

$$Homogeneity = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2} \quad (13)$$

4. *Angular Second Moment (ASM)*
ASM digunakan untuk menghitung

seberapa tinggi tingkat keteraturan piksel.

$$ASM = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{i,j}^2 \quad (14)$$

5. *Entropy*
Entropy digunakan untuk mengukur tingkat ketidakaturan piksel.

$$Entropy = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{i,j} (-\ln P_{i,j}) \quad (15)$$

6. *GLCM Mean vertical*
Rata-rata GLCM yang dihitung dengan banyaknya kemunculan kombinasi dikali piksel tetangga.

$$\mu_i = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} i(P_{i,j}) \quad (16)$$

7. *GLCM Mean horizontal*
 $\mu_j = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} j(P_{i,j}) \quad (17)$

8. *Variance vertical*
Pengukuran untuk mengetahui penyebaran nilai disekitar rata-rata.

$$\sigma_i = \sqrt{\sum_{i,j=0}^{N-1} (P_{i,j})(i - \mu_i)^2} \quad (18)$$

9. *Variance horizontal*
 $\sigma_j = \sqrt{\sum_{i,j=0}^{N-1} (P_{i,j})(j - \mu_j)^2} \quad (19)$

10. *Correlation*
Correlation mengukur ketergantungan linear tingkat keabuan pada piksel dengan piksel tetangganya.

$$Correlation = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{i,j} \frac{(i - \mu_i)(j - \mu_j)}{\sqrt{\sigma_i^2 \sigma_j^2}} \quad (20)$$

Keterangan:

P_{ij} = nilai *GLCM* pada index ke i dan j

N = ukuran matrik *GLCM*

μ_i = *GLCM mean horizontal*

μ_j = *GLCM mean vertical*

σ_i = *variance horizontal*

σ_j = *variance vertical*

2.6 Information Gain

Information Gain merupakan salah satu metode pemilihan fitur berdasarkan nilai *ranking* atribut yang banyak diaplikasikan pada analisis data citra, kategorisasi teks, dan analisis microarray (Chormunge & Jena, 2016).

$$Entropy(S) = \sum_i^c -P_i \log_2 P_i \quad (21)$$

$$Gain(S, A) = Entropy(S) - \sum_{values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (22)$$

Di mana c adalah jumlah nilai yang ada pada kelas klasifikasi, P_i adalah jumlah sampel untuk kelas i, A adalah atribut, v adalah nilai yang mungkin untuk atribut A, $values(A)$ adalah himpunan nilai-nilai yang mungkin untuk A, $|S_v|$ adalah jumlah sampel untuk nilai v, $|S|$ adalah jumlah seluruh sampel data, dan $Entropy(S_v)$ adalah nilai *entropy* masing-masing sampel.

Namun karena data yang dipakai pada penelitian ini bersifat kontinu maka harus dikategorisasikan terlebih dahulu dengan menggunakan distribusi frekuensi (Darwiyanto,

Binawan dan Junaedi, 2017). Langkah-langkahnya adalah sebagai berikut.

1. Menentukan banyak kelas dengan persamaan 22.
2. menentukan panjang setiap data dengan persamaan 23

$$kelas = 1 + 3.3\log(n_k) \quad (23)$$

$$jangkauan = (\max(x) - \min(x))/kelas \quad (24)$$

Di mana n_k adalah jumlah kelas data yang digunakan. X adalah data fitur yang sama.

2.7 Dissimilarity Measure

Perhitungan kesamaan pada penelitian ini menggunakan 2 perhitungan jarak yang berbeda-beda untuk tiap fitur. Kemudian dijumlahkan dengan dikali bobot untuk masing-masing fitur seperti yang dilakukan oleh (Wang, Yu, & Yang, 2011).

$$D(DCD, GLCM) = w_{DCD}D_{DCD} + w_{GLCM}D_{GLCM} \quad (25)$$

Di mana $w_{DCD} + w_{GLCM} = 1$. w_{DCD} adalah bobot untuk fitur warna. w_{GLCM} adalah bobot untuk fitur tekstur.

1. Chebyshev Distance

Chebyshev distance merupakan perhitungan jarak diantara 2 vektor atau 2 titik p dan q dengan standar koordinat p_i dan q_i . *Chebyshev distance* merupakan perhitungan jarak turunan dari *Minkowski distance* dengan $p = \infty$ (Pasumarthi & Malleswari, 2016).

$$D_{Chebyshev}(p, q) = \max_i(|p_i - q_i|) \quad (26)$$

Di mana i adalah fitur pada data, q adalah data pertama, dan p adalah data kedua.

2. Euclidean Distance

Euclidean distance merupakan perhitungan jarak turunan dari *Minkowski distance* dengan $p = 2$ (Pasumarthi & Malleswari, 2016).

$$D_{Euclidean}(i, j) = \sqrt{\sum_{k=1}^n (x_k(i) - x_k(j))^2} \quad (27)$$

Di mana x_k adalah atribut ke k , n merupakan banyak atribut, i merupakan data pertama, j merupakan data kedua.

3. Manhattan Distance

Manhattan distance merupakan perhitungan jarak turunan dari *Minkowski distance* dengan $p = 1$ (Pasumarthi & Malleswari, 2016).

$$D_{Manhattan}(i, j) = \sum_{k=1}^n |x_k(i) - x_k(j)| \quad (28)$$

Di mana x_k adalah atribut ke k , n merupakan banyak atribut, i merupakan data pertama, j merupakan data kedua.

4. Quadratic Distance

Quadratic distance didasarkan pada adanya satu nilai koherensi spasial yang mewakili keseluruhan citra dan beberapa kelompok dari (*centroid*, presentase) untuk warna dominan yang sesuai.

$$D(F_1, F_2)$$

$$= \sqrt{\sum_{i=1}^{N_1} P_{1i}^2 + \sum_{j=1}^{N_2} P_{2j}^2 - \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} 2a_{1i,2j}P_{1i}P_{2j}} \quad (29)$$

$$a_{1i,2j} = \begin{cases} 1 - \frac{d_{1i,2j}}{d_{max}} & d_{1i,2j} \leq T_d \\ 0 & d_{1i,2j} > T_d \end{cases} \quad (30)$$

$$d_{1i,2j} = \|C_{1i} - C_{2j}\| \quad (31)$$

$$d_{max} = \alpha T_d \quad (32)$$

Keterangan:

$F_1 = DCD$ citra 1

$F_2 = DCD$ citra 2

P_{1i} = Presentase warna indeks i pada citra 1

P_{2j} = Presentase warna indeks j pada citra 2

N_1 = Jumlah *descriptor* pada citra 1

N_2 = Jumlah *descriptor* pada citra 2

C_{1i} = Titik *centroid* indeks i pada citra 1

C_{2j} = Titik *centroid* indeks j pada citra 2

α = Parameter yang bernilai (1, 1.5)

T_d = Batasan jarak maksimum yang digunakan untuk menentukan apakah 2 warna mirip atau tidak.

5. Yang Distance

Perhitungan jarak yang diusulkan oleh (Yang, et al., 2008) yang memberikan konsistensi pada persepsi manusia dan meningkatkan performa perhitungan.

$$D_{yang}(I_1, I_2) = 1 - Sim_{yang}(I_1, I_2) \quad (33)$$

$$Sim_{yang}(I_1, I_2) = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} a_{1i,2j} S_{1i,2j} \quad (34)$$

$$S_{1i,2j} = [1 - |P_{1i} - P_{2j}|] \times \min(P_{1i}, P_{2j}) \quad (35)$$

Keterangan:

$a_{1i,2j}$ = Kesamaan warna seperti pada persamaan 30

P_{1i} = Presentase warna indeks i pada citra 1

P_{2j} = Presentase warna indeks j pada citra 2

2.8 Mean Average Precision (MAP)

Mean Average Precision adalah evaluasi temu kembali informasi yang menyediakan satu nilai pengukuran dari kualitas untuk semua tingkat *recall*. *MAP* mempunyai kelebihan pada diskriminasi dan stabilitas (Manning, 2009).

$$MAP(Q) = \frac{1}{|Q|} \sum_{j=1}^{|Q|} \frac{1}{m_j} \sum_{k=1}^{m_j} P(R_{jk}) \quad (36)$$

Keterangan:

$|Q|$ = Banyaknya kueri

M_j = Banyaknya dokumen relevan untuk kueri j

$P(R_{jk})$ = *Precision* pada dokumen relevan ke k pada kueri j

3. HASIL DAN PEMBAHASAN

3.1 Pengujian Standard Deviasi Fitur

Pengujian ini bertujuan untuk mengetahui nilai *standard deviasi* terhadap fitur setelah

dilakukan seleksi fitur. Hasil pengujian ini dapat dilihat pada Tabel 2.

Pada hasil pengujian nilai *standard deviasi*, nilai *standard deviasi* pada 10 fitur teratas hasil seleksi fitur *information gain* memiliki nilai-nilai yang kecil dibawah 1. Namun fitur-fitur dengan urutan 11-20 memiliki nilai-nilai yang besar yaitu sekitar 20. Kemudian pada fitur-fitur dengan urutan 21-30 memiliki nilai-nilai yang kecil kembali yaitu sekitar 2. Kemudian pada fitur-fitur dengan urutan 31-40 memiliki nilai-nilai yang kecil lagi yaitu dibawah sekitar 0,1. Sehingga dapat dikatakan hasil seleksi fitur *information gain* tidak memiliki pengaruh dengan hasil pengujian nilai *standard deviasi*.

3.2 Pengujian Menggunakan Fitur Tekstur

Pengujian ini bertujuan untuk melihat pengaruh perhitungan jarak pada fitur tekstur. Hasil dari pengujian ini dapat dilihat pada Tabel 1.

Tabel 1 Hasil Pengujian Fitur Tekstur

Jarak	10 Fitur	20 Fitur	30 Fitur	Tanpa Seleksi
Chebyshev	0,4366	0,3963	0,5393	0,5393
Euclidean	0,4508	0,3890	0,5519	0,5542
Manhattan	0,4463	0,3918	0,5321	0,5435

Pada Tabel 1 dapat diketahui nilai MAP tertinggi didapatkan ketika tidak menggunakan seleksi fitur. Namun pada penggunaan fitur sebanyak 30, nilai MAP yang didapatkan memiliki jarak yang sangat dekat dengan ketika tidak memakai seleksi fitur. Dari hasil pengujian tekstur ini yang disajikan oleh Tabel 6.4, diketahui nilai MAP terendah yaitu ketika pengujian menggunakan perhitungan jarak *euclidean distance* dengan menggunakan 20 fitur yaitu sebesar 0,389 dan nilai MAP tertinggi didapatkan ketika pengujian menggunakan perhitungan jarak *euclidean distance* tanpa menggunakan seleksi fitur yaitu sebesar 0,5542.

Tabel 2 Hasil Pengujian Standard Deviasi Fitur

No.	Fitur	Standar Deviasi
1	glcm_90_homogeneity	0.043270704
2	glcm_45_dissimilarity	0.148190963
3	glcm_135_dissimilarity	0.148190963
4	glcm_0_homogeneity	0.041226362
5	glcm_45_homogeneity	0.051212376
6	glcm_135_homogeneity	0.051212376
7	glcm_90_entropy	0.511567315
8	glcm_90_dissimilarity	0.117833761
9	glcm_45_entropy	0.541533008

10	glcm_135_entropy	0.541533008
11	glcm_0_entropy	0.507914787
12	glcm_0_dissimilarity	0.1126748
13	glcm_0_variance_baris	20.14693218
14	glcm_0_variance_kolom	20.14693218
15	glcm_45_variance_baris	20.15681317
16	glcm_45_variance_kolom	20.15681317
17	glcm_90_variance_baris	20.14586187
18	glcm_90_variance_kolom	20.14586187
19	glcm_135_variance_baris	20.15681317
20	glcm_135_variance_kolom	20.15681317
21	glcm_0_mean_baris	2.522128346
22	glcm_0_mean_kolom	2.522128346
23	glcm_45_mean_baris	2.522780838
24	glcm_45_mean_kolom	2.522780838
25	glcm_90_mean_baris	2.522427289
26	glcm_90_mean_kolom	2.522427289
27	glcm_135_mean_baris	2.522780838
28	glcm_135_mean_kolom	2.522780838
29	glcm_45_contrast	0.523774616
30	glcm_135_contrast	0.523774616
31	glcm_90_contrast	0.364380451
32	glcm_45_correlation	0.009408767
33	glcm_135_correlation	0.009408767
34	glcm_90_correlation	0.00732834
35	glcm_0_contrast	0.356958045
36	glcm_0_correlation	0.006429532
37	glcm_45_asm	0.061922909
38	glcm_135_asm	0.061922909
39	glcm_0_asm	0.061848001
40	glcm_90_asm	0.061795149
27	glcm_135_mean_baris	2.522780838
28	glcm_135_mean_kolom	2.522780838
29	glcm_45_contrast	0.523774616
30	glcm_135_contrast	0.523774616
31	glcm_90_contrast	0.364380451
32	glcm_45_correlation	0.009408767
33	glcm_135_correlation	0.009408767
34	glcm_90_correlation	0.00732834
35	glcm_0_contrast	0.356958045
36	glcm_0_correlation	0.006429532
37	glcm_45_asm	0.061922909
38	glcm_135_asm	0.061922909
39	glcm_0_asm	0.061848001
40	glcm_90_asm	0.061795149

3.3 Pengujian Menggunakan Fitur Warna

Pengujian ini dilakukan untuk mengetahui pengaruh perhitungan jarak pada fitur warna. Hasil dari pengujian ini disajikan pada Tabel 3.

Tabel 3 Hasil Penujian Fitur Warna

Jarak	MAP
Yang	0,7488
Quadratic	0,6813

Berdasarkan hasil pengujian menggunakan fitur warna yang disajikan pada Tabel 3 dapat diketahui bahwa nilai MAP tertinggi ketika menggunakan perhitungan jarak *yang distance*

yaitu sebesar 0,7488 dan nilai MAP terendah ketika menggunakan perhitungan jarak *quadratic distance* yaitu sebesar 0,6813. Selisih antara nilai MAP tertinggi dan nilai MAP terendah cukup kecil yaitu sebesar 0,0675. Sehingga walaupun perhitungan jarak menggunakan *Yang distance* lebih unggul, namun tidak memiliki jarak yang signifikan untuk mengungguli perhitungan jarak menggunakan *quadratic distance*. Dari Tabel 3 juga dapat dilihat bahwa penggunaan fitur warna lebih baik dari pada penggunaan fitur tekstur untuk temu kembali citra jenis makanan ini.

3.4 Pengujian Menggunakan Fitur Tekstur dan Fitur Warna

Pengujian ini dilakukan untuk mengetahui pengaruh penggunaan fitur tekstur dan fitur warna dengan menggunakan kombinasi perhitungan jarak yang berbeda. Hasil dari pengujian ini dapat dilihat pada Tabel 4.

Tabel 4 Hasil Pengujian Menggunakan Fitur Tekstur dan Fitur Warna

Jarak	10 Fitur	20 Fitur	30 Fitur	Tanpa Seleksi
chebyshev yang	0,6984	0,6975	0,6953	0,6954
chebyshev quadratic	0,6936	0,6701	0,6656	0,6688
euclidean yang	0,7046	0,6975	0,6967	0,6975
euclidean quadratic	0,6707	0,6704	0,6684	0,6702
manhattan yang	0,7118	0,6964	0,7053	0,7049
manhattan quadratic	0,6729	0,6712	0,6800	0,6822

Berdasarkan pada hasil pengujian menggunakan fitur tekstur dan fitur warna yang disajikan pada Tabel 4, dapat diketahui nilai MAP terendah pada saat menggunakan perhitungan jarak *chebyshev distance* dan *quadratic distance* pada saat menggunakan 30 fitur hasil seleksi fitur yaitu sebesar 0,6656. Sedangkan nilai MAP terbesar didapatkan ketika menggunakan perhitungan jarak menggunakan *manhattan distance* dan *yang distance* dan menggunakan 10 fitur hasil seleksi fitur yaitu sebesar 0,7118.

4. KESIMPULAN

Berdasarkan hasil pengujian dan pembahasan, dapat diambil kesimpulan sebagai berikut:

Dari hasil pengujian hanya menggunakan

fitur tekstur saja memiliki nilai MAP tertinggi yaitu sebesar 0,5542 dengan menggunakan jarak *euclidean*. Namun nilai terbesar pada penggunaan fitur tekstur saja masih lebih rendah dari pada hanya menggunakan fitur warna saja yang memiliki nilai MAP tertinggi sebesar 0,7488 dengan menggunakan jarak *Yang*. Namun dari hasil pengujian yang menggunakan fitur tekstur dan fitur warna memiliki hasil nilai MAP yang meningkat dari hanya menggunakan fitur tekstur yaitu sebesar 0,7118. Walaupun masih belum mengungguli hasil pengujian menggunakan fitur warna saja, sudah memiliki selisih yang sedikit. Dari hasil pengujian diketahui penggunaan seleksi fitur *information gain* hanya bekerja pada penggunaan fitur tekstur dan fitur warna di mana mendapatkan hasil yang lebih baik dengan menggunakan 10 fitur dengan perhitungan jarak *chebyshev* dan *Yang*.

Pada penelitian ini masih diperlukan penelitian mengenai metode ekstraksi fitur yang digunakan karena pada penggunaan metode GLCM ini masih didapat hasil yang lebih rendah dari penggunaan fitur warna.

5. DAFTAR PUSTAKA

- Aini, S.H.A., Sari, Y.A. dan Arwan, A., 2018. Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(9), pp.2546–2554.
- Bahmani, B., Moseley, B., Vattani, A., Kumar, R. dan Vassilvitskii, S., 2012. Scalable K-Means++. [online] Tersedia di: <<http://arxiv.org/abs/1203.6402>>.
- Chormunge, S. dan Jena, S., 2016. Efficient Feature Subset Selection Algorithm for High Dimensional Data. *International Journal of Electrical and Computer Engineering (IJECE)*, [online] 6(4), p.1880. Tersedia di: <<http://www.iaescore.com/journals/index.php/IJECE/article/view/758>>.
- Darwiyanto, E., Binawan, B.P. dan Junaedi, D., 2017. Aplikasi GIS Klasifikasi Tingkat Kerawanan Banjir Wilayah Kabupaten Bandung Menggunakan Metode Weighted Product. 2, pp.59–70.
- Hall-Beyer, M., 2017. Gldm Texture: a Tutorial.

- International Journal of Remote Sensing*, [online] 38(5), pp.1312–1338. Tersedia di: <<https://www.tandfonline.com/doi/full/10.1080/01431161.2016.1278314>>.
- Li, A. & Bao, X., 2010. Extracting image dominant color features based on region growing. *Proceedings - 2010 International Conference on Web Information Systems and Mining, WISM 2010*, 2(07), pp.120–123.
- OpenCV, 2016. Color Conversions. [online] Tersedia di: <https://docs.opencv.org/3.2.0/de/d25/imgproc_color_conversions.html> [Diakses pada 16 Desember 2018]
- Pasumarthi, N. dan Malleswari, L., 2016. An empirical study and comparative analysis of Content Based Image Retrieval (CBIR) techniques with various similarity measures. *3rd International Conference on Electrical, Electronics, Engineering Trends, Communication, Optimization and Sciences (EEECOS 2016)*, [online] 3, pp.1–6. Tersedia di: <<https://ieeexplore.ieee.org/document/8372206>>.
- Rasli, R.M., Muda, T.Z.T., Yusof, Y. dan Bakar, J.A., 2012. Comparative Analysis of Content Based Image Retrieval Techniques Using Color Histogram: A Case Study of GLCM and K-Means Clustering. *2012 Third International Conference on Intelligent Systems Modelling and Simulation*, [online] pp.283–286. Tersedia di: <<http://ieeexplore.ieee.org/document/6169714/>>.
- Sharma, M. dan Batra, A., 2014. Analysis of Distance Measures in Content Based Image Retrieval. 14(2).
- Tobey, L.N. dan Manore, M.M., 2014. Viewpoint Social Media and Nutrition Education: The Food Hero Experience. *Journal of Nutrition Education and Behavior*, [online] 46(2), pp.128–133. Tersedia di: <<http://dx.doi.org/10.1016/j.jneb.2013.09.013>>.
- Verdaguer, S.L., 2009. Color Based Image Classification and Description.
- Wang, X.Y., Yu, Y.J. dan Yang, H.Y., 2011. An effective image retrieval scheme using color, texture and shape features. *Computer Standards and Interfaces*, [online] 33(1), pp.59–68. Tersedia di: <<http://dx.doi.org/10.1016/j.csi.2010.03.004>>.
- Yang, N., Kuo, C., Chang, W. dan Lee, T., 2008. A Fast Method for Dominant Color Descriptor with New Similarity Measure. *Iscom2005*. [online] Tersedia di: <http://140.117.166.1/eehome/ISCOM2005/SubmitPaper/UploadPapers/ISCON05_00089.pdf>.