

BAB III

LANDASAN TEORI

3.1 Citra Sawit

Citra merupakan suatu fungsi intensitas cahaya dua dimensi $f(x,y)$ dimana x dan y menyatakan suatu koordinat spasial, nilai f pada $f(x,y)$ sebanding dengan dari citra tersebut. (Gustina et al., 2017). 3 Jenis sawit yang menjadi objek penelitian terdiri dari sawit jenis dura, pisifera dan tenera. Pada umumnya jenis dura memiliki ciri ciri daging buah relatif tipis, yaitu 35-50% terhadap buah, jenis pisifera memiliki ciri ciri fisik ketebalan tempurung sangat tipis, bahkan hampir tidak ada sedang pada sawit jenis tenera daging buah sangat tebal. Contoh citra buah sawit yang digunakan pada penelitian ini seperti pada gambar 3.1 Sawit jenis Dura, 3.2 Sawit Jenis Pisifera dan 3.3 Sawit jenis Tenera.



Sumber Gambar: blog.tokotanaman.com

Gambar 3. 1 Sawit Jenis Dura

Daging buah relatif tipis, yaitu 35-50% terhadap buah, serta tempurung buah tebal, tandan buah lebih banyak, tetapi ukurannya relatif kecil



Sumber Gambar: blog.tokotanaman.com

Gambar 3. 2 Sawit Jenis Pisifare

Ketebalan tempurung sangat tipis, bahkan hampir tidak ada. Daging buah tebal, lebih tebal dari daging buah Dura.



Sumber: blog.tokotanaman.com

Gambar 3. 3 Sawit Jenis Tenera.

Daging buah sangat tebal (60-96% dari buah) dan ukuran tempurung tipis. Sehingga menghasilkan minyak yang lebih banyak, Perbedaan per varian dapat dilihat pada dibawah ini.

Tabel 3. 1 Perbandingan varietas sawit

Varietas	Deskripsi
Dura	Tempurung tebal (2-8 mm) Daging buah relatif tipis, yaitu 35-50% terhadap buah
Pisifera	Ketebalan tempurung sangat tipis, bahkan hampir tidak ada Daging buah tebal, lebih tebal dari daging buah Dura
Tenera	Tempurung tipis (0,5-4 mm) Daging buah sangat tebal (60-96% dari buah) Tandan buah lebih banyak, tetapi ukurannya relatif kecil

Sumber: (Fauzi, Widyastuti, Satyawibawa, & Paeru, 2012)

3.2 Segmentasi Citra

Segmentasi citra merupakan tahapan penting dalam proses preprocessing data. Setelah objek tersegmentasi dengan benar, maka kita dapat melakukan proses ekstraksi ciri citra. Ekstraksi ciri adalah langkah yang bertujuan untuk mengekstraksi fitur dari suatu objek di mana fitur tersebut digunakan untuk membedakan antara objek satu dengan objek lainnya. Thresholding adalah metode atau cara yang paling sederhana dari segmentasi citra. Nilai threshold dapat digunakan dari citra grayscale untuk menghasilkan citra biner. Citra biner adalah citra digital yang hanya memiliki dua kemungkinan nilai untuk setiap pikselnya. Kedua warna tersebut adalah hitam dan putih dengan piksel objek menjadi 1 sedangkan latar belakangnya adalah 0. (Bhahri & Rachmat, 2018), maka dilakukan konversi dengan mengambil rata-rata dari nilai r, g dan b:

$$S_{segmentation} = (r+g+b)/3$$

Citra abu-abu merupakan citra dengan nilai intensitas pikselnya didasarkan pada derajat keabuan. Pada citra abu-abu 8-bit, derajat warna hitam sampai dengan putih dibagi ke dalam 256 derajat keabuan di mana warna hitam sempurna direpresentasikan dengan nilai 0 dan putih sempurna dengan nilai 255. Citra RGB dapat dikonversi menjadi citra grayscale sehingga dihasilkan hanya satu kanal warna.

3.3 Pengolahan Citra Digital (*Digital Image Processing*)

Pengolahan citra digital atau *digital image processing* secara umum didefinisikan sebagai proses komputerisasi pada citra dua dimensi. Menurut Zonyfar (2020, hal. 2-3) proses pengolahan pada citra perlu dilakukan apabila:

- a. Perlu adanya perbaikan atau modifikasi citra untuk meningkatkan kualitas penampakan atau untuk menonjolkan suatu aspek tertentu.
- b. Terdapat kecacatan pada citra sehingga perlu untuk dihilangkan.
- c. Perlu pengelompokan terhadap elemen tertentu.
- d. Perlu ekstraksi ciri-ciri tertentu misal tekstur, warna, dan bentuk.
- e. Perlu penggabungan dengan citra lain.
- f. Menyembunyikan informasi rahasia.

Tujuan dari pengolahan citra digital adalah membantu manusia untuk meningkatkan kualitas atau mendeskripsikan karakteristik suatu citra. Proses pendefinisian terhadap citra tersebut dibutuhkan sebuah implementasi algoritma kecerdasan untuk mendeteksi objek dalam citra.

Pada penelitian ini akan dilakukan ekstraksi terhadap citra untuk mendapatkan data terkait warna, tekstur, dan bentuk tertentu.

3.4 Metode Red Green Blue (RGB)

RGB merupakan salah satu format penyimpanan citra yang dibuat oleh *silicon graphics* untuk menyimpan citra berwarna. Citra berwarna (*truecolor*) adalah citra yang memiliki tiga kanal warna di dalamnya, citra ini dimodelkan ke dalam ruang warna dengan R (*red/merah*), G (*green/hijau*), dan B (*Blue/biru*) sebagai pembentuk

komponen citra. Pada layer computer biasanya menggunakan RGB sebagai standar warna dalam menampilkan citra berwarna (Zonyfar, 2020).

3.5 Algoritma Naïve Bayes

Naïve Bayes Classifier merupakan salah satu metode pengklasifikasian yang berdasarkan teorema Bayes. Metode ini menggunakan probabilitas dan statistik yang digunakan untuk memprediksi peluang di masa depan berdasarkan pada kejadian-kejadian sebelumnya. Naïve Bayes Classifier ini memiliki ciri utama yaitu memiliki asumsi yang kuat akan independensi dari setiap kondisi atau kejadian. Terdapat tiga Naïve Bayes Classifier yang populer yaitu, Gaussian Naïve Bayes Classifier, Multinomial Naïve Bayes Classifier dan Bernoulli Naïve Bayes Classifier.

Naïve Bayes Classifier memiliki tingkat Keakuratan lebih baik dibanding dengan model pengklasifikasian lainnya. Penelitian yang dilakukan oleh (Singh, Kumar, & Gaur, 2019) terhadap pengklasifikasian text (*Text Classification*) untuk membantu mempermudah *Machine Learning* menggunakan metode multinomial dan Naïve Bayes. Hasil penelitian ini Naïve Bayes memiliki hasil yang lebih akurat. Penelitian yang dilakukan oleh (Xhemali, Hinde, & Stone, 2009) untuk mengklasifikasikan *training web page* menggunakan Naïve Bayes, Decision Tree, dan Neural Network. Kesimpulan penelitian ini, dari tiga metode pengklasifikasian, metode yang memberikan hasil paling akurat adalah metode Naïve Bayes.

3.6 Hue Saturasi Value (HSV)

Pemodel warna HSV mendefinisikan warna dalam terminologi Hue, Saturation dan Value. Hue menyatakan warna sebenarnya, seperti merah, violet, dan kuning. Hue digunakan untuk membedakan warna-warna dan menentukan kemerahan (redness), kehijauan (greeness), dsb, dari cahaya. Hue berasosiasi dengan panjang gelombang cahaya. Saturation menyatakan tingkat kemurnian suatu warna, yaitu mengindikasikan seberapa banyak warna putih diberikan pada warna. Value adalah atribut yang menyatakan banyaknya cahaya yang diterima oleh mata tanpa memperdulikan warna (Purnamasari et al., 2013)

3.7 Klasifikasi Naïve Bayes

Salah satu metode yang menggunakan konsep probabilitas adalah metode naive bayes. Pada metode ini, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain. Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Langkah-langkah untuk pelatihan data.

- 1) Hitung probabilitas (prior) tiap kelas yang ada.
- 2) Lalu hitung rata-rata (mean) tiap fitur dan tiap kelas.
- 3) Hitung nilai standar deviasi tiap fitur dan tiap kelas.
- 4) Hitung densitas probabilitas.
- 5) hitung probabilitas masing-masing kelas.

Berikut rumus persamaan dari teorema bayes (Bustami, 2014):

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Dimana :

X = Data dengan kelas yang belum diketahui

H = Hipotesis data merupakan suatu class spesifik

$P(H|X)$ = Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

$P(H)$ = Probabilitas hipotesis H (prior probabilitas)

$P(X|H)$ = Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$ = Probabilitas X

3.8 Confusion Matrix

Confusion matrix adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining dan Supervised Learning. Pada pengukuran kinerja menggunakan confusion matrix, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi, keempat istilah tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Nilai *True Negative* (TN) merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Positive* (FP) merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, *True Positive* (TP) merupakan data positif yang terdeteksi benar. *False Negative* (FN) merupakan kebalikan dari *True Positive*,

sehingga data positif, namun terdeteksi sebagai data negatif (Rokach & Maimon, 2015) .

Presisi adalah data yang diambil berdasarkan informasi yang kurang. Dalam klasifikasi biner, presisi dapat dibuat sama dengan nilai prediksi positif. Berikut ini adalah aturan presisi.

$$\text{Precision} = (\text{TP} / (\text{TP} + \text{FP})) * 100\%$$

Recall adalah data penghapusan yang berhasil diambil dari data yang relevan dengan kueri. Dalam klasifikasi *biner*, *recall* dikenal sebagai sensitivitas. Munculnya data relevan yang diambil adalah menyetujui dengan query dapat dilihat dengan recall. Berikut ini adalah peran recall

$$\text{Recall} = (\text{TP} / (\text{TP} + \text{FN})) * 100\%$$

Akurasi adalah persentase dari total data yang diidentifikasi dan dinilai. Berikut ini adalah aturan akurasi

$$\text{Akurasi} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) * 100\%$$

3.9 Extraksi Ciri Warna

Untuk membedakan objek dengan warna tertentu, terlebih dahulu harus dapat menggunakan nilai hue yang merupakan representasi cahaya tampak (merah, jingga, kuning, hijau, biru, ungu). Nilai rona dapat digabungkan dengan nilai

saturasi, dan value adalah tingkat kecerahan warna. Untuk mendapatkan ketiga nilai ini, perlu mengkonversi ruang warna gambar dari RGB (merah, hijau, biru) ke HSV (hue, saturation, value) menggunakan rumus berikut :

$$\begin{aligned}
 R' &= R/255 \\
 G' &= G/255 \\
 B' &= B/255 \\
 C_{max} &= \max(R', G', B') \\
 C_{min} &= \min(R', G', B') \\
 \Delta &= C_{max} - C_{min}
 \end{aligned}$$

Perhitungan Hue

$$H = \begin{cases} 0^\circ & \Delta = 0 \\ 60^\circ \times \left(\frac{G' - B'}{\Delta} \text{ mod } 6 \right) & , C_{max} = R' \\ 60^\circ \times \left(\frac{B' - R'}{\Delta} + 2 \right) & , C_{max} = G' \\ 60^\circ \times \left(\frac{R' - G'}{\Delta} + 4 \right) & , C_{max} = B' \end{cases}$$

Perhitungan Saturasi

$$S = \begin{cases} 0 & , C_{max} = 0 \\ \frac{\Delta}{C_{max}} & , C_{max} \neq 0 \end{cases}$$

Perhitungan Value

$$V = C_{max}$$

Perhitungan grayscale:

$$\text{Grayscale} = 0,2989 \cdot R + 0,5870 \cdot G + 0,1140 \cdot B$$

3.10 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) adalah salah satu metode manifold learning yang merupakan suatu metode untuk mengekstrak fitur dengan cara mereduksi dimensi dari dimensi tinggi ke dimensi yang lebih rendah (Anggo & La Arapu.) . Berikut ini adalah tahapantahapan dari metode LDA

1. Tahapan pertama pada proses LDA adalah mengubah data citra latih dan citra uji menjadi vektor-vektor

$$\{x_1, x_2, x_3 \dots x_n\}$$

2. Buat kelas-kelas berdasarkan banyaknya data pada data citra latih dan citra uji.
3. Hitung rata-rata dalam kelas (m_i) dan rata-rata keseluruhan kelas (m) dari seluruh citra di database
4. Menghitung matriks sebaran antar kelas (S_b) menggunakan persamaan

$$\sum_{i=1}^k n_i (m_i - m)(m_i - m)^t$$

5. Menghitung matriks sebaran dalam kelas menggunakan persamaan
6. Mencari matriks W dengan menggunakan persamaan
7. Mencari vektor eigen dari W dan mengurutkan nilai eigen (λ) sesuai dengan urutan nilai yang ada pada nilai eigen dari besar ke kecil.
8. Melakukan reduksi dimensi dengan cara melakukan proyeksi dengan vektor eigen yang sudah dipilih ke ruang dimensi yang lebih rendah yang disebut dengan fisher basis vector menggunakan persamaan.
9. Dimana : adalah vektor hasil proyeksi yang berdimensi lebih rendah