



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB II

LANDASAN TEORI

2.1 Pepaya

Pepaya adalah tanaman yang diduga berasal dari Meksiko bagian selatan hingga Nikaragua, dan masuk ke Indonesia pada abad ke-17 (Hamzah, 2014). Dalam ilmu botani, pepaya diklasifikasikan sebagai berikut.

1. *Kingdom* : *Plantae* (Tumbuhan)
2. *Subkingdom* : *Tracheobionta* (Tumbuhan berpembuluh)
3. *Super Divisi* : *Spermatophyta* (Tumbuhan berbiji)
4. *Divisi* : *Magnoliophyta* (Tumbuhan berbunga)
5. *Kelas* : *Magnoliopsida* (Berkeping dua/dikotil)
6. *Subkelas* : *Dilleniidae*
7. *Ordo* : *Violales*
8. *Familia* : *Caricaceae*
9. *Genus* : *Carica*
10. *Spesies* : *Carica pepayae* L.

Menurut Khomsan (2009), pepaya dapat tumbuh di setiap tempat, dari pesisir, pegunungan, hingga pekarangan rumah. Beberapa khasiat yang dimiliki oleh buah pepaya yaitu mengatasi gangguan pencernaan, pembesaran hati dan limpa, penyakit kulit, alergi, serta menguatkan lambung (Adrian dan Dalimartha, 2011). Rasa buah pepaya akan semakin manis saat matang. Jika sudah matang, berbagai jenis pepaya berubah warna kulitnya menjadi kuning atau jingga kemerahan, dari yang awalnya berwarna hijau saat masih mentah (Suryobuwono dkk., 2005).

2.2 Semangka

Semangka adalah tanaman yang berasal dari Afrika, yang sekarang telah menyebar luas di daerah subtropik maupun tropic (Rukmana, 1994). Dalam ilmu botani, pepaya diklasifikasikan sebagai berikut (Sobir dan Siregar, 2010).

1. *Kingdom* : *Plantae* (Tumbuhan)
2. *Divisi* : *Magnoliophyta* (Tumbuhan berbunga)
3. *Kelas* : *Magnoliopsida* (Berkeping dua/dikotil)
4. *Ordo* : *Violales*
5. *Familia* : *Cucurbitaceae*
6. *Genus* : *Citrullus*
7. *Spesies* : *Citrullus lanatus*

Sobir dan Siregar (2010) juga menyatakan bahwa tanaman semangka dapat ditanam pada daerah-daerah yang kurang dapat ditanami dengan komoditas lain karena kebutuhannya akan air yang relatif rendah. Menurut Adrian dan Dalimartha (2011), beberapa manfaat buah semangka bagi kesehatan tubuh yaitu mengurangi risiko kanker dan penyakit jantung koroner, menjaga tekanan darah tetap normal, serta menurunkan kadar asam urat. Buah semangka yang sudah matang dapat dilihat pada bagian ujung, jika warnanya putih kehijauan, berarti masih belum matang. Akan tetapi, jika warnanya sudah merata dengan warna pada bagian lain dan terlihat mengilap, maka artinya sudah matang.

2.3 Satuan Brix

Menurut Wulan (2010), satuan *brix* adalah satuan yang digunakan untuk mengukur kandungan padatan terlarut dalam suatu larutan. Pada buah, kandungan

padatan terlarut terdiri atas gula-gula sederhana, seperti fruktosa, glukosa, dan sukrosa. Semakin tinggi nilai derajat *brix*, maka semakin manis larutan tersebut. Besar larutan yang diambil adalah 100 gram. Jika diperoleh nilai derajat *brix* 12%, maka 12 gram larutan tersebut adalah kandungan padatan terlarut, sedangkan 88 gram merupakan air.

Faktor cuaca dapat mempengaruhi tingkat kemanisan buah. Curah hujan yang tinggi akan menurunkan rasa manis buah (Redaksi Trubus, 2007). Oleh karena itu, buah akan terasa lebih manis pada musim kemarau dibandingkan pada musim hujan.

Klasifikasi tingkat kemanisan buah pepaya dan semangka dapat dilihat pada Tabel 2.1.

Tabel 2.1 Tingkat kemanisan buah pepaya dan semangka (Harrill, 1998)

| Buah | Tidak Manis | Cukup Manis | Manis | Sangat Manis |
|----------|-------------|-------------|-------|--------------|
| Pepaya | 6 | 10 | 18 | 22 |
| Semangka | 4 | 6 | 8 | 12 |

2.4 Pengolahan Citra Digital

Pengolahan citra digital adalah pemrosesan gambar dua dimensi menggunakan komputer (Putra, 2010). Tujuan pengolahan citra digital ada dua, yaitu memperbaiki kualitas gambar, dan mengolah informasi yang terdapat di dalamnya untuk mengenali pola (Pramesti dkk., Tanpa Tahun). Dalam penelitian ini, tujuannya yaitu untuk mengenali pola. Pola yang dimaksud adalah nilai komponen warna *red*, *green*, dan *blue* (RGB) tiap *pixel* yang membentuk citra. Nilai tersebut dapat diketahui dengan menggunakan rumus persamaan normalisasi sebagai berikut.

$$r = \frac{R}{R+G+B} \quad (2.1)$$

$$g = \frac{G}{R+G+B} \quad (2.2)$$

$$b = \frac{B}{R+G+B} \quad (2.3)$$

Setelah itu, nilai RGB akan digunakan dalam mengidentifikasi tingkat kemanisan buah.

2.5 Segmentasi

Untuk memisahkan antara objek yang ingin diolah dan latar belakangnya, dilakukan segmentasi. Salah satu tekniknya adalah *thresholding*. Dengan proses ini, citra *grayscale* akan berubah menjadi citra biner, yaitu citra dengan warna hitam dan putih (Nugroho, 2012). Proses *thresholding* dapat dirumuskan sebagai berikut (Putra, 2010).

$$g(x,y) = \begin{cases} 1 & \text{if } f(x,y) \geq T \\ 0 & \text{if } f(x,y) < T \end{cases} \quad (2.4)$$

Keterangan :

$g(x,y)$: Citra biner

$f(x,y)$: Citra *grayscale*

T : Nilai ambang

Secara umum, *thresholding* ada dua jenis, yaitu *global thresholding* dan *locally adaptive thresholding*. Pada *global thresholding*, seluruh *pixel* dikonversi dengan satu nilai T sehingga dapat mengakibatkan banyak informasi yang hilang, sedangkan pada *locally adaptive thresholding*, citra dibagi menjadi blok-blok kecil, dengan masing-masing blok memiliki nilai T yang berbeda. Menurut Putra

(2010), kualitas hasil citra bergantung pada nilai T . Oleh karena itu, *locally adaptive thresholding* secara umum lebih baik.

2.6 Naïve Bayes Classifier

Naïve Bayes Classifier adalah metode pengklasifikasian sederhana yang menggunakan konsep probabilitas (Pramesiti dkk., Tanpa Tahun). Algoritma ini dikemukakan oleh Thomas Bayes, dan digunakan untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa lalu (Bustami, 2014). Menurut Pattekari dan Parveen (2012), keuntungan penggunaan klasifikasi ini yaitu hanya membutuhkan data pelatihan (*training data*) yang sedikit. Secara sederhana, rumus *naive Bayes classifier* dapat ditulis sebagai berikut (Saleh, 2015).

$$\mathbf{Posterior} = \frac{\mathbf{prior} \times \mathbf{likelihood}}{\mathbf{evidence}} \quad (2.5)$$

Prior adalah peluang muncul suatu kelas sebelum sampel masuk. Peluang merupakan perbandingan antara banyak hasil yang dimaksud dan banyak hasil yang mungkin keluar (Untoro, 2007). Oleh karena itu, *prior* dapat dihitung menggunakan persamaan sebagai berikut.

$$\mathbf{Peluang} = \frac{\mathbf{banyak\ hasil\ yang\ dimaksud}}{\mathbf{banyak\ hasil\ yang\ mungkin\ keluar}} \quad (2.6)$$

Nilai *prior* kemudian dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas tersebut (*likelihood*), lalu dibagi dengan kemunculan karakteristik-karakteristik sampel secara global (*evidence*). Oleh karena sifatnya global, maka nilai *evidence* akan selalu sama untuk setiap kelas.

Untuk menghitung salah satu karakteristik sampel pada suatu kelas dapat digunakan rumus berikut (Saleh, 2015).

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad (2.7)$$

- Keterangan :
- P : Peluang
 - X_i : Atribut ke- i
 - x_i : Nilai atribut ke- i
 - Y : Kelas yang dicari
 - y_j : Subkelas Y yang dicari
 - μ : *Mean*, menyatakan rata-rata dari seluruh atribut
 - σ : Deviasi standar, menyatakan varian dari seluruh atribut.

Evidence dapat disebut juga sebagai *normalizing constant*. Secara umum, *normalizing constant* adalah total peluang untuk mendapatkan data yang diteliti (MacKay, 2004). Jika terdapat dua kelas, rumus untuk menghitung *normalizing constant* dijabarkan sebagai berikut.

$$P(s | F) = P(s | F_1 H_1) \times P(H_1) + P(s | F_0 H_0) \times P(H_0) \quad (2.8)$$

Keterangan :

- $P(s | F)$: *Evidence*
- $P(s | F_1 H_1)$: *Likelihood* kelas ke-1
- $P(H_1)$: *Prior* kelas ke-1
- $P(s | F_0 H_0)$: *Likelihood* kelas ke-2

$P(H_0)$: *Prior* kelas ke-2

Nilai dari *posterior* masing-masing kelas suatu sampel kemudian akan dibandingkan untuk mengidentifikasi kelas sampel tersebut. Kelas yang memiliki nilai *posterior* paling tinggi berarti adalah kelas dari sampel.

2.7 Overfitting

Salah satu teknik untuk menilai performa dari suatu sistem adalah dengan melakukan uji coba terhadap data yang dimiliki. Data tersebut dapat kemudian dibagi menjadi dua, yaitu *training data* yang membentuk sistem, dan *test data* yang akan diuji (Maindonald dan Braun, 2010).

Menurut Rokach dan Maimon (2008), *overfitting* adalah situasi di mana algoritma klasifikasi yang digunakan hanya baik untuk data yang sama persis dengan data yang digunakan sebagai *training*, sedangkan saat digunakan data yang baru untuk *testing*, kapabilitasnya berkurang. Kapabilitas dalam penelitian ini yaitu tingkat akurasi sistem. Dengan kata lain, sistem bukan melakukan pembelajaran, melainkan menghafal *training data*.

U
M
N