

SEGMENTASI DAUN GALBUN MENGGUNAKAN ALGORITMA C4.5 DAN NAIVE BAYES

Hamdun Sulaiman

Universitas Bina Sarana Informatika
e-mail: hamdun.hsl@bsi.ac.id

Diterima	Direvisi	Disetujui
15-12-2019	16-12-2019	17-12-2019

Abstrak - Segmentasi adalah proses mempartisi citra digital menjadi beberapa bagian. Tujuan dari segmentasi yaitu untuk menyederhanakan atau mengubah penyajian gambar ke sesuatu yang lebih bermakna dan lebih mudah untuk dianalisa dalam menentukan nilai dari suatu gambar. Segmentasi citra biasanya digunakan untuk menemukan obyek dan batas-batas dalam sebuah citra. Penelitian ini menggunakan sebuah citra daun galbun, dimana daun galbun tersebut dibagi menjadi dua kelas yaitu daun galbun sehat dan daun galbun sakit. Dalam prosesnya digunakan dengan cara men segmentasi citra daun galbun tersebut menggunakan *Software* MATLAB untuk mendapatkan nilai dari setiap segmentasi daun galbun, dan nilai hasil dari segmentasi di olah ke dalam *Data Mining* menggunakan *Rapid Miner* dengan perbandingan algoritma C4.5 dan Naive Bayes. Dari hasil penelitian ini di dapat bahwa algoritma C4.5 menghasilkan akurasi sebesar 90%, sedangkan Naive Bayes mendapatkan akurasi 95%. Jadi algoritma yang cocok atau terbaik untuk menghitung akurasi dari hasil segmentasi tersebut yaitu algoritma Naive Bayes dengan nilai 95%.

Kata Kunci: segmentasi, citra, data mining

Abstract - Segmentation is the process of partitioning digital images into parts. The purpose of segmentation is to simplify or change the presentation of images to something more meaningful and easier to analyze in determining the value of an image. Image segmentation is usually used to find objects and boundaries in an image. This study uses a galbun leaf image, where the galbun leaf is divided into two classes, namely healthy galbun leaf and sick galbun leaf. In the process used by segmenting the galbun leaf image using MATLAB Software to get the value of each galbun leaf segmentation, and the resulting value of the segmentation is processed into Data Mining using Rapid Miner with a comparison of C4.5 and Naive Bayes algorithms. From the results of this study it was found that the C4.5 algorithm produced an accuracy of 90%, while Naive Bayes gained an accuracy of 95%. So the best or best algorithm for calculating the accuracy of the results of the segmentation is the Naive Bayes algorithm with a value of 95%.

Keywords: segmentation, image, data mining

PENDAHULUAN

Pengolahan citra adalah istilah umum untuk berbagai teknik yang keberadaannya untuk memanipulasi dan memodifikasi citra dengan berbagai cara (Permata, 2016). Citra merupakan istilah lain untuk gambar sebagai salah satu komponen multimedia yang memegang peranan yang sangat penting sebagai bentuk informasi visual. Citra mempunyai karakteristik yang tidak dimiliki oleh data teks, yaitu citra kaya dengan informasi (Permadi & Murinto, 2015).

Menurut (Permadi & Murinto, 2015) Citra (image) secara harfiah adalah gambar pada bidang dwimatra (dua dimensi). Ditinjau dari sudut pandang matematis, citra merupakan fungsi menerus (continue) dari intensitas cahaya pada bidang dwimatra. Sumber cahaya menerangi objek, objek memantulkan kembali sebagai dari berkas cahaya

tersebut. Pantulan cahaya ini ditangkap oleh alat-alat optik, misalnya mata pada manusia, kamera, scanner, dan sebagainya. Sehingga bayangan objek yang disebut citra tersebut terakam.

Dalam visi komputer, segmentasi adalah proses mempartisi citra digital menjadi beberapa bagian. Tujuan dari segmentasi adalah untuk menyederhanakan atau mengubah penyajian gambar ke sesuatu yang lebih bermakna dan lebih mudah untuk dianalisa. Gambar segmentasi biasanya digunakan untuk menemukan obyek dan batas-batas dalam sebuah citra. Segmentasi citra merupakan proses menempatkan label untuk setiap piksel dalam sebuah gambar sehingga piksel dengan kumpulan label yang sama karakteristik visual tertentu (Permata, 2016).

(Permata, 2016) mengatakan bahwa hasil segmentasi citra adalah seperangkat segmentasi yang secara kolektif mencakup seluruh gambar atau satu

set kontur yang diekstrak dari citra. Setiap piksel dalam suatu wilayah memperhatikan beberapa karakteristik properti, seperti warna, intensitas dan tekstur. Daerah yang berdekatan sangat berbeda sehubungan dengan karakteristik yang sama.

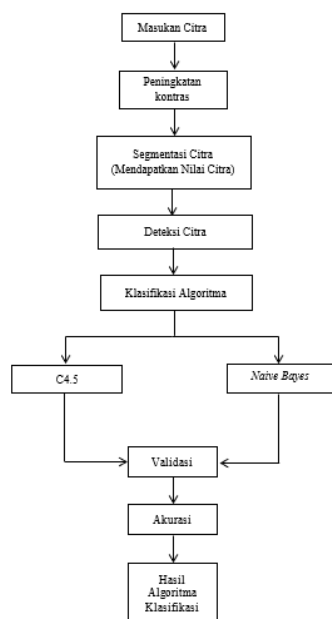
Algoritma Klasifikasi *Data Mining* adalah suatu metode pembelajaran untuk memprediksi nilai dari sekelompok atribut dalam menggambarkan dan membedakan kelas data atau konsep yang bertujuan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Beberapa algoritma klasifikasi yang sering digunakan antara lain naive bayes, decision tree, neural network, k-nn, random forest dan lain sebagainya. Performa algoritma data mining dalam banyak kasus tergantung pada kualitas dataset, karena data training berkualitas rendah dapat menyebabkan klasifikasi yang lemah (Ardiyansyah, Rahayuningsih, & Maulana, 2018).

Dalam penelitian ini akan membandingkan algoritma decision tree C4.5 dan naive bayes pada hasil segmentasi menggunakan citra daun galbun, yang nantinya algoritma mana yang cocok atau yang mempunyai nilai akurasi terbaik untuk hasil dari segmentasi tersebut.

METODE PENELITIAN

Penelitian ini akan membandingkan algoritma C4.5 dan *Naive Bayes* untuk mencari mana algoritma yang terbaik pada hasil segmentasi citra daun galbun.

Pengujian terhadap metode yang akan digunakan, membutuhkan desain penelitian yang jelas dan tepat. Desain yang akan dibuat akan digunakan sebagai tahapan proses pengujian untuk mendapatkan hasil yang diinginkan. Maka dari itu pada Gambar 1 menjelaskan sebuah model desain penelitian yang dilakukan.



Gambar 1 Model Desain Penelitian

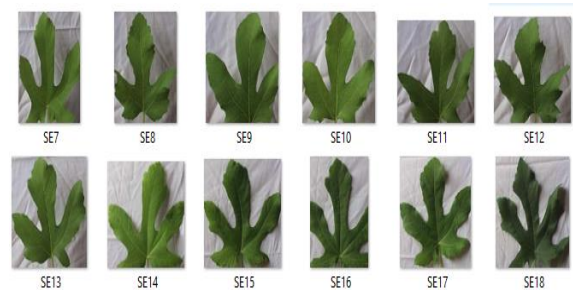
HASIL DAN PEMBAHASAN

A. Hasil Penelitian

Penelitian ini menghasilkan keluaran yang dapat dianalisa untuk mendapatkan informasi yang berguna. Berikut penjabaran dari hasil penelitian yang sudah dilakukan .

B. Pengumpulan Data

Dalam penelitian ini metode pengumpulan data untuk mendapatkan sumber data yang digunakan adalah metode pengumpulan data primer. Data utama citra daun galbun diperoleh dari hasil pengambilan gambar sendiri dengan beberapa citra daun, sedangkan data pendukung didapatkan dari buku, jurnal dan publikasi lainnya. Dan akan men segmentasi citra daun galbun dengan menggunakan daun galbun sehat dan sakit masing-masing 30 citra.



Sumber: (Yosep Petani Galbun, 2019)
Gambar 2. Citra Daun Sehat



Sumber: (Yosep Petani Galbun, 2019)
Gambar 3. Citra Daun Sakit

C. Segmentasi Citra Daun Galbun

Di bawah ini merupakan hasil dari proses segmentasi citra daun galbun dalam GUI MATLAB. Pertama masukan citra daun yang akan di segmentasi, kemudian meningkatkan kontrasnya, selanjutnya di segmentasi citra yang dimasukan, yang nantinya akan ada clustering untuk memilih hasil segmentasi dan terakhir ada deteksi dari setiap citra daun yang di segmentasi apakah termasuk daun sehat atau sakit.



Sumber: (pengolahan segmentasi matlab)
Gambar 4. Proses Segmentasi

D. Pengujian Algoritma

Pengujian Algoritma bertujuan membandingkan algoritma C4.5 dan algoritma *Naive Bayes* pada nilai hasil segmentasi dengan citra daun galbun. Yang nantinya algoritma mana yang memiliki nilai akurasi terbaik

1. Hasil Pengujian Algoritma C4.5

Algoritma C4.5 diperkenalkan oleh *J. Ross Quinlan* yang merupakan pengembangan dari algoritma ID3, algoritma tersebut digunakan untuk membentuk pohon keputusan. Pohon keputusan dianggap sebagai salah satu pendekatan yang paling populer. Dalam klasifikasi pohon keputusan terdiri dari sebuah node yang membentuk akar. Node akar tidak memiliki input. Node lain yang bukan sebagai akar tetapi memiliki tepat satu input disebut node internal atau test node, sedangkan node lainnya dinamakan daun. Daun mewakili nilai target yang paling tepat dari salah satu class (Mukminin & Riana, 2017).

Pada dasarnya konsep dari algoritma C4.5 adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan (rule). C4.5 adalah algoritma yang cocok untuk masalah klasifikasi dan data mining. C4.5 memetakan nilai atribut menjadi class yang dapat diterapkan untuk klasifikasi baru (Mukminin & Riana, 2017).

a. Evaluasi dengan *Confusion Matrix*

Confusion Matrix adalah suatu metode untuk melakukan evaluasi dengan menggunakan tabel matrix yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining (Rosandy, 2016). Evaluasi dengan menggunakan fungsi confusion matrix akan menghasilkan nilai *accuracy*, *precision*, dan *recall* (Buani, 2016).

Setelah data uji dimasukkan kedalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *accuracy*, *sensitivity* untuk mengukur proporsi ‘*positif*’ yang benar yang didiagnosa dengan benar, *specificity* untuk mengukur proporsi ‘*negatif*’ yang benar yang didiagnosa dengan benar, *PPV (Positive Predictive Value)* adalah proporsi kasus dengan hasil tes ‘*positif*’ yang didiagnosa dengan benar, dan *NPV (Negative Predictive Value)* adalah proporsi kasus

dengan hasil tes ‘*negatif*’ yang didiagnosa dengan benar. Dapat dihitung menggunakan rumus (Gorunescu, 2011, p. 320):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Sensitivity = \frac{TP}{TP+FN}$$

$$Specificity = \frac{TN}{TN+FP}$$

$$PPV = \frac{TP}{TP+FP}$$

$$NPV = \frac{TN}{TN+FN}$$

Keterangan:

TP : jumlah true positives

TN : jumlah true negatives

FP : jumlah false positives

FN : jumlah false negatives

Model *confusion matrix* akan menghasilkan *matrix* yang terdiri dari *true positif* atau *true negative* setelah data testing di masukan kedalam model *confusion matrix* hasilnya akan terlihat seperti gambar dibawah:

accuracy 90.00% +/- 15.20% (jumlah: 90.00%)			
	true Daun Sehat	true Daun Sakit	class precision
pred. Daun Sehat	28	4	87.50%
pred. Daun Sakit	2	26	92.86%
class recall	93.33%	86.67%	

Sumber: (proses pengujian *rapid miner*)

Gambar 5. *Confusion Matrix* Algoritma C4.5

Berdasarkan Gambar 5 dari 60 set data, bahwa algoritma C4.5 dapat memprediksi kelas Daun Sehat sebanyak 28 dan 4 kelas Daun Sehat dianggap Daun Sakit. Kelas Daun Sakit dapat diprediksi sebanyak 26 dan 2 kelas Daun Sakit dianggap Daun Sehat. Sehingga didapatkan class precision atau sensitivity untuk memprediksi kelas Daun Sehat sebesar 87,50% dan untuk kelas Daun Sakit atau specificity sebesar 92,86%. Class recall untuk true Daun Sehat atau PPV sebesar 93,33% dan true Daun Sakit atau NPV sebesar 86,67%.

Dari data tersebut maka dapat dihitung nilai *accuracy* dengan menggunakan persamaan.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Accuracy = \frac{28+26}{28+26+4+2} = 0,90 \text{ (90\%)}$$

2. Hasil Pengujian Algoritma *Naive Bayes*

Naive Bayes merupakan metode yang tidak memiliki aturan. *Naive Bayes* menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada data training. *Naive Bayes* merupakan metode klasifikasi populer dan masuk dalam sepuluh algoritma terbaik dalam data mining. Algoritma ini juga dikenal dengan nama *Idiot's*

Bayes, Simple Bayes, dan Independence Bayes (Mukminin & Riana, 2017).

Menurut (Mukminin & Riana, 2017) Klasifikasi *Naive Bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Klasifikasi Bayesian didasarkan pada teorema Bayes, diambil dari nama seorang ahli matematika yang juga menteri Prebysterian Inggris, *Thomas Bayes* (1702-1761). Menurut Kusri dan Luthfi, klasifikasi bayesian memiliki kemampuan klasifikasi serupa dengan decision tree dan *neural network*. *Bayes rule* digunakan untuk menghitung probabilitas suatu class. Algoritma *Naive Bayes* memberikan suatu cara mengkombinasikan peluang terdahulu dengan syarat kemungkinan menjadi sebuah formula yang dapat digunakan untuk menghitung peluang dari tiap kemungkinan yang terjadi.

a. Evaluasi dengan *Confusion Matrix*

Confusion Matrix adalah suatu metode untuk melakukan evaluasi dengan menggunakan tabel matrix yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining (Rosandy, 2016). Evaluasi dengan menggunakan fungsi confusion matrix akan menghasilkan nilai *accuracy*, *precision*, dan *recall* (Buani, 2016).

Setelah data uji dimasukkan kedalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *accuracy*, *sensitivity* untuk mengukur proporsi ‘positif’ yang benar yang didiagnosa dengan benar, *specificity* untuk mengukur proporsi ‘negatif’ yang benar yang didiagnosa dengan benar, *PPV (Positive Predictive Value)* adalah proporsi kasus dengan hasil tes ‘positif’ yang didiagnosa dengan benar, dan *NPV (Negative Predictive Value)* adalah proporsi kasus dengan hasil tes ‘negatif’ yang didiagnosa dengan benar. Dapat dihitung menggunakan rumus (Gorunescu, 2011, p. 320):

$$\begin{aligned} Accuracy &= (TP+TN)/(TP+TN+FP+FN) \\ Sensitivity &= TP/(TP+FN) \\ Specificity &= TN/(TN+FP) \\ PPV &= TP/(TP+FP) \\ NPV &= TN/(TN+FN) \end{aligned}$$

Keterangan:

- TP : jumlah true positives
- TN : jumlah true negatives
- FP : jumlah false positives
- FN : jumlah false negatives

Model *confusion matrix* akan menghasilkan matrix yang terdiri dari *true positif* atau *true negative* setelah *testing* di masukan kedalam model *confusion matrix* hasilnya akan terlihat seperti gambar dibawah:

accuracy: 95.00% +/- 10.67% (nilai: 95.00%)			
	true Daun Sehat	true Daun Sakit	class precision
pred. Daun Sehat	30	3	90.91%
pred. Daun Sakit	0	27	100.00%
class recall	100.00%	90.00%	

Sumber: (proses pengujian *rapid miner*)
Gambar 6.

Confusion Matrix Algoritma *Naive Bayes*

Berdasarkan Gambar 4.5. dari 60 set data, bahwa algoritma *naive bayes* dapat memprediksi kelas Daun Sehat sebanyak 30 dan 3 kelas Daun Sehat dianggap Daun Sakit. Kelas Daun Sakit dapat diprediksi sebanyak 27 dan 0 kelas Daun Sakit dianggap Daun Sehat. Sehingga didapatkan class precision atau sensitivity untuk memprediksi kelas Daun Sehat sebesar 90,91% dan untuk kelas Daun Sakit atau specificity sebesar 100%. Class recall untuk true Daun Sehat atau PPV sebesar 100% dan true Daun Sakit atau NPV sebesar 90%.

Dari data tersebut maka dapat dihitung nilai *accuracy* dengan menggunakan persamaan.

$$\begin{aligned} Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} \\ &= \frac{30+27}{30+27+0+3} = 0,95 \text{ (95\%)} \end{aligned}$$

KESIMPULAN

Dari hasil penelitian untuk nilai terbaik pada perbandingan algoritma dengan hasil dari segmentasi citra daun galbun. Untuk nilai akurasi terhadap algoritma klasifikasi C4.5 sebesar 90%, sedangkan untuk nilai akurasi algoritma *Naive Bayes* sebesar 95%

Dapat disimpulkan bahwa algoritma *Naive Bayes* memiliki nilai terbaik dalam hasil segmentasi citra tersebut.

REFERENSI

Ardiyansyah, Rahayuningsih, P. A., & Maulana, R. (2018). Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner. *Jurnal Khatulistiwa Informatika*, VI(1), 20–28. <https://doi.org/10.31294/jki.v6i1.3799.g2437>

Buani. (2016). Optimasi Algoritma *Naive Bayes* dengan Menggunakan Algoritma Genetika untuk Prediksi Kesuburan (Fertility). *Evolusi*, 4(1), 55–64. <https://doi.org/10.2311/evo.v4i1.277>

Mukminin, A., & Riana, D. (2017). Komparasi Algoritma C4.5, *Naive Bayes* Dan Neural Network Untuk Klasifikasi Tanah. *Jurnal Informatika*, 4(1), 21–31.

- Permadi, Y., & Murinto, . (2015). Aplikasi Pengolahan Citra Untuk Identifikasi Kematangan Mentimun Berdasarkan Tekstur Kulit Buah Menggunakan Metode Ekstraksi Ciri Statistik. *Jurnal Informatika*, 9(1), 1028–1038. <https://doi.org/10.26555/jifo.v9i1.a2044>
- Permata, E. (2016). Identifikasi Obyek Benda Tajam Menggunakan Pengolahan Citra Digital Pada Citra X-Ray. *Volt*, 1(1), 1–14. <https://doi.org/10.1124/jpet.102.042010.Capsa> zepine
- Rosandy, T. (2016). PERBANDINGAN METODE NAIVE BAYES CLASSIFIER DENGAN METODE DECISION TREE (C4.5) UNTUK MENGANALISA KELANCARAN PEMBIAYAAN (Study Kasus: KSPPS / BMT AL-FADHILA. *Jurnal Teknologi Informasi Magister Darmajaya*, 2(1), 52–62.