

## PENERAPAN ALGORITMA C4.5 PADA BIDANG PERTANIAN

Aris Haris Rismayana<sup>1)</sup>, Dian Rosdiana<sup>2)</sup>

<sup>1),2)</sup> Teknik Informatika, Politeknik TEDC

Email: rismayana@poltektedc.ac.id<sup>1)</sup>, dianrosdiana72@gmail.com<sup>2)</sup>

### Abstrak

Budidaya cabai selalu menarik untuk didiskusikan, karena cabai termasuk salah satu bahan dasar yang selalu terdapat dalam makanan. Ketika harga tinggi, petani cabai akan mendapatkan keuntungan sangat besar. Namun sebaliknya ketika harga cabai anjlok, petani sangat bingung sehingga mereka enggan merawat cabai mereka lagi. Fluktuasi harga cabai sangat dipengaruhi oleh cuaca, panen, kebijakan perdagangan dan momentum yang menyertainya. Tindak lanjut dari penanaman ketika harga tinggi juga mengubah struktur harga. Jika orang menanam hanya karena motif mengikuti harga cabai yang tinggi saat ini, maka seorang petani tidak akan mampu bertahan, mengingat fluktuasi harga cabai sangat tinggi. Berdasarkan fenomena tersebut, maka salah satu tujuan dari penelitian ini adalah untuk Memprediksi waktu tanam cabai menggunakan algoritma C4.5. Proses pengumpulan data cuaca dan harga dilakukan oleh peneliti yang diambil dari DISPERINDAG dan BMKG serta data ketersediaan cabai di pasar untuk digunakan sebagai data training. Data training dikumpulkan melalui proses preprocessing dengan melakukan kategorisasi, kemudian dilakukan training untuk menghasilkan model. Dari model kemudian dibuat aplikasi yang mampu memprediksi waktu tanam cabai. Hasil akurasi yang didapat sebesar 72.2 % dengan kriteria akurasi Good classification menggunakan confusion matrix.

**Kata kunci:** waktu tanam cabai, harga cabai, data mining, klasifikasi, C4.5, prediksi

### Abstract

*Chili cultivation is always interesting to discuss, because chili is one of the basic ingredients that is always found in food. When prices are high, chili farmers will benefit greatly. But on the contrary when the price of chili dropped, farmers were so confused that they were reluctant to care for their chili again. Fluctuations in the price of chilies are greatly influenced by weather, harvest, trade policies and the accompanying momentum. Follow-up from planting when prices are high also changes the price structure. If people plant only because the motive is to follow the current high prices of chili, then a farmer will not be able to survive, given the fluctuations in chili prices are very high. Based on this phenomenon, then one of the objectives of this study is to predict the planting time of chili using the C4.5 algorithm. The process of collecting weather and price data is carried out by researchers drawn from DISPERINDAG and BMKG as well as chili availability data on the market to be used as training data. Training data is collected through the process of preparing by categorizing, then training is conducted to produce a model. Then an application is made from the model that is able to predict the planting time of chili. Accuracy results obtained by 72.2% with Good classification accuracy criteria using a confusion matrix.*

**Keywords:** chili planting time, chili prices, data mining, classification, C4.5, prediction

## I. PENDAHULUAN

Budidaya cabai selalu menarik untuk didiskusikan. Ketika harga tinggi, petani akan menerima keuntungan sangat besar. Namun sebaliknya ketika anjlok harga cabai, petani bingung untuk tidak merawat cabai mereka lagi. Fluktuasi harga cabai sangat dipengaruhi oleh cuaca, musim panen, kebijakan perdagangan dan momentum yang menyertainya. Tindak lanjut dari penanaman ketika harga tinggi juga mengubah struktur harga.

Jika ingin menanam selama musim kemarau atau sekitar Juni hingga Agustus, ada kemungkinan petani akan mendapatkan harga rendah di awal panen, tetapi masih bisa mendapatkan harga menengah dan harga tinggi di musim hujan dengan syarat bahwa tanaman masih terawat dengan baik dan panen masih optimal sampai akhir panen. Agar hasilnya tetap optimal,

meski penanaman di musim kemarau, ketinggian gundukan harus tetap 40 cm untuk mengantisipasi masuknya musim hujan. Selain itu, aplikasi pestisida biologis juga perlu diberikan untuk menjaga kondisi tanaman.

## II. LANDASAN TEORI

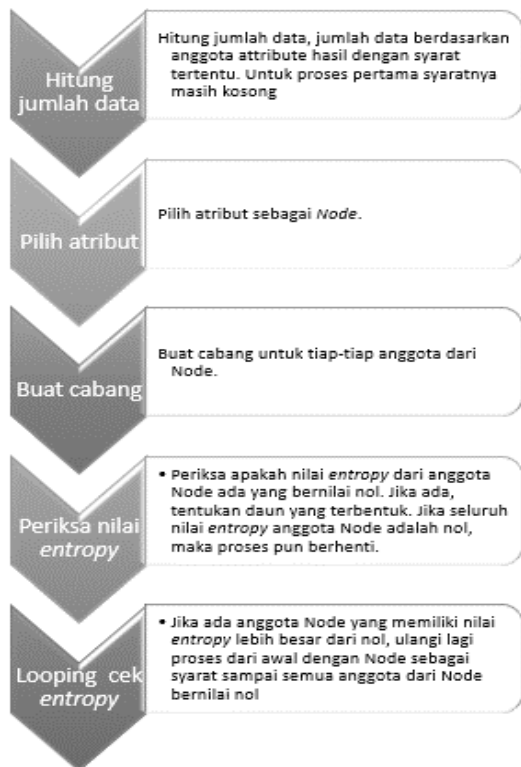
*Data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. *Algoritma decision tree* didasarkan pada pendekatan *divide-and-conquer* untuk klasifikasi suatu masalah.

Algoritma tersebut bekerja dari atas ke bawah, mencari pada setiap tahap atribut untuk membaginya ke dalam bagian terbaik dari class tersebut dan memproses secara rekursif submasalah yang dihasilkan dari pembagian

tersebut. Strategi ini menghasilkan sebuah decision tree yang dapat diubah menjadi satu set classification rules Kenneth (2004). Decision tree merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (tree) di mana setiap node merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas.

Sebelum algoritma C4.5, yang sedang berkembang saat itu adalah algoritma ID3. Baik Algoritma C4.5 ataupun algoritma ID3, sama-sama diciptakan oleh J. Rose Quinlan, seorang peneliti di bidang Artificial Intelegent, sekitar akhir tahun 1970 Gonzales (2004). Algoritma C4.5 membuat keputusan secara sekuensial, berurutan dari atas ke bawah. dimana atribut paling atas merupakan akar, dan yang paling bawah dinamakan daun.

Secara umum, algoritma C4.5 untuk membangun sebuah pohon keputusan adalah sebagai berikut:



Gambar 1. Bagan decision tree C4.5

Untuk menghitung nilai gain suatu atribut digunakan rumus seperti yang tertera dalam persamaan berikut:

$$Gain(S, A) = Entropy(S) - \left[ \sum_{i=1}^n \frac{A_i}{S} * Entropy(A_i) \right] \quad (1)$$

Keterangan:

- S : Kasus.
- A : Atribut
- N : Jumlah partisi atribut A
- A<sub>i</sub> : Jumlah kasus pada partisi ke-i.
- S : Jumlah kasus.

Sementara itu, untuk menghitung nilai Entropy dapat dilihat pada persamaan berikut ini:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

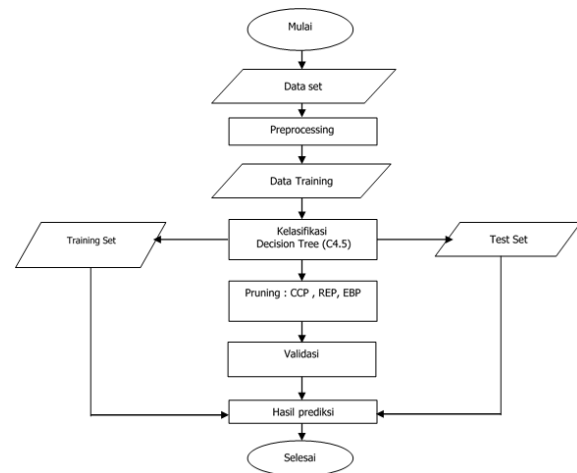
Keterangan:

- S : Himpunan kasus.
- N : Jumlah partisi S
- p<sub>i</sub> : Proporsi dari S<sub>i</sub> ke S.

### III. METODE PENELITIAN

Pada penelitian ini digunakan dua metode untuk dua aktivitas utama yaitu dokumentasi dan implementasi.

Dibawah ini adalah alur penelitian:



Gambar 2. Blok diagram sistem

Dari diagram blok sistem dapat dijelaskan sebagai berikut:

- a. Data mentah
 

Data yang digunakan berupa data historis yang diperoleh dari Departemen Perindustrian dan Perdagangan (DISPERINDAG) dan data curah hujan yang digunakan sebagai data pendukung diperoleh dari Badan Meteorologi, Klimatologi dan Geofisika (BMKG).
- b. Preprocessing
 

Tahap ini dilakukan untuk memudahkan algoritma pembacaan input data yang akan diproses. Data harga dan curah hujan merupakan input dari sistem yang akan dibangun. Dalam penelitian ini, data akan diproses melalui pembersihan data dan transformasi data. Pembersihan data menghilangkan atribut atau data yang diperoleh tetapi tidak diperlukan dalam penelitian ini, sedangkan transformasi data adalah dengan mengkategorikan data yang telah diperoleh.
- c. Data pelatihan
 

Setelah fase preprocessing selesai, data ini siap untuk dijadikan data pelatihan, data pelatihan diambil untuk diproses ke dalam

- algoritma C4.5 untuk menemukan model prediksi waktu tanam cabai yang tepat.
- d. Proses pohon keputusan (C4.5)
 

Metode ini adalah salah satu metode yang tersedia dalam teknik klasifikasi dalam penambangan data. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang mewakili aturan.
  - e. Proses pemangkasan: CCP, REP, EBP
 

Pemangkasan adalah bagian dari proses pembentukan pohon keputusan. Saat membentuk pohon keputusan, beberapa node adalah outlier dan hasil dari data noise. Penerapan pemangkasan ke pohon keputusan dapat mengurangi data pencilaan dan kebisingan di pohon keputusan awal sehingga dapat meningkatkan akurasi klasifikasi data (Han & Kamber 2006). Pemangkasan yang tepat perlu dilakukan untuk mendapatkan hasil klasifikasi maksimum. Dalam studi yang dilakukan oleh (Esposito et al. 1997), algoritma Reduced Error Pruning (REP) disimpulkan sebagai algoritma yang menghasilkan subtree terkecil dengan tingkat kesalahan minimum. Penelitian Esposito et al. (1997) menggunakan algoritma C4.5 untuk membangun pohon keputusan pemangkasan. Algoritma CART yang biasa menggunakan *Cost Complexity Pruning* (CCP) sebagai algoritma pemangkasannya. Dalam penelitian ini algoritma pemangkasan CCP dibandingkan dengan dua algoritma pemangkasan lainnya, REP dan Error Based Pruning (EBP). Penelitian ini bertujuan untuk menerapkan teknik CCP, REP dan EBP untuk metode klasifikasi pohon keputusan menggunakan algoritma CART. Selain itu, penelitian ini juga membandingkan nilai akurasi dari pohon keputusan yang dibentuk, dan waktu pemrosesan yang dihasilkan oleh algoritma pemangkasan CCP, REP dan EBP.
  - f. Validasi
 

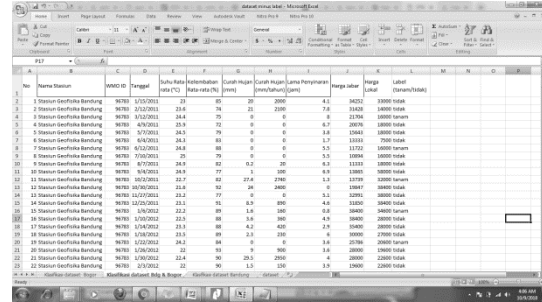
Pengujian dilakukan dengan menggunakan teknik validasi silang k-flip. Metode validasi silang digunakan untuk memprediksi akurasi data pengujian (Hastie & Tibshirani, 2009).
  - g. Hasil prediksi
 

Output yang dihasilkan oleh algoritma C4.5

**IV. HASIL DAN PEMBAHASAN**

**Proses Pengolahan Data**

Data didapat dari Dinas Perindustrian dan Perdagangan (DISPERINDAG) serta data curah hujan yang dijadikan data pendukung didapatkan dari Badan Meteorologi, Klimatologi dan Geofisika (BMKG). Data yang dikumpulkan yaitu data harga komoditi cabai provinsi jawa barat dan cuaca sebanyak 858 dari tahun 2011 sd 2018 dengan attribute Suhu rata-rata, kelembaan rata-rata, Curah hujan, harga.



**Gambar 3.** Dataset mentah untuk proses

Setelah data dimasukkan atau diolah menggunakan *Microsoft Office Excel*, maka dilakukan pengolahan data dengan cara mengambil data rata-rata x1 (Suhu), rata-rata x2 (Kelembaban), rata-rata x3 (Curah Hujan), rata-rata x4 (Harga) dan Y (Hasil) dan dilakukan pemilihan atribut pada proses pengolahan dibawah ini:

No	Tanggal	Suhu (X1)	Kelembaban (X2)	Curah Hujan	Harga Lokal (X4)	Hasil (Y)	Label (tanam/tidak)
1	1/15/2011	23	85	20	33000	Baik	tidak
2	2/12/2011	23.6	74	21	14000	Tidak	tidak
3	3/12/2011	24.4	75	0	16000	Tidak	tanam
4	4/9/2011	25.9	72	0	18000	Tidak	tidak
5	5/7/2011	24.5	79	0	18000	Tidak	tidak
6	6/4/2011	24.3	83	0	7500	Tidak	tidak
7	6/12/2011	24.8	88	0	16000	Tidak	tanam
8	7/10/2011	25	79	0	16000	Tidak	tidak
9	8/7/2011	24.9	82	0.2	18000	Baik	tidak
10	9/4/2011	24.9	77	1	58000	Baik	tidak
11	10/2/2011	22.7	82	27.4	32000	Baik	tanam
12	10/30/2011	21.6	92	24	38400	Baik	tidak
13	11/27/2011	23.2	77	0	38000	Baik	tidak
14	12/25/2011	23.1	91	8.9	38400	Baik	tidak
15	1/6/2012	22.2	89	1.6	34600	Baik	tanam
16	1/10/2012	22.5	88	3.6	28000	Baik	tidak
17	1/14/2012	23.3	88	4.2	28000	Baik	tidak
18	1/18/2012	23.5	89	2.3	27000	Baik	tidak
19	1/22/2012	24.2	84	0	20600	Baik	tanam
20	1/26/2012	22	93	9	19600	Tidak	tidak
21	1/30/2012	22.4	90	29.5	22600	Baik	tidak
22	2/3/2012	22	90	1.5	22600	Baik	tidak

**Gambar 4.** Dataset setelah preprocessing

**Proses Data Mining**

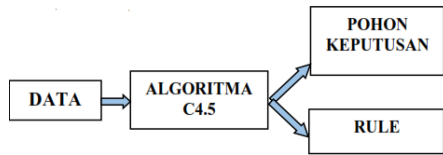
Pada tahap selanjutnya data akan di proses menggunakan bantuan *Tools Rapid Miner* dan untuk pengujianya menggunakan cross validation yang ada di dalam *Rapid Miner*, maka data akan terlihat nilai akurasi, pohon keputusan, dan rulenya. Berikut adalah

Data yang akan di import ke dalam *Rapid Miner*:

Tanggal	Suhu	Kelembaban	Curah Hujan	Harga Lokal	Hasil	Label (tanam/tidak)
Polynomial	Numeric	Numeric	Numeric	Numeric	Bionomial	Bionomial
Id	Attribute	Attribute	Attribute	Attribute	Label	Label
1/15/2011	23	85	20	33000	Baik	tidak
2/12/2011	23.6	74	21	14000	Tidak	tidak
3/12/2011	24.4	75	0	16000	Tidak	tanam
4/9/2011	25.9	72	0	18000	Tidak	tidak
5/7/2011	24.5	79	0	18000	Tidak	tidak
6/4/2011	24.3	83	0	7500	Tidak	tidak
6/12/2011	24.8	88	0	16000	Tidak	tanam
7/10/2011	25	79	0	16000	Tidak	tidak
8/7/2011	24.9	82	0.2	18000	Baik	tidak
9/4/2011	24.9	77	1	58000	Baik	tidak
10/2/2011	22.7	82	27.4	32000	Baik	tanam
10/30/2011	21.6	92	24	38400	Baik	tidak
11/27/2011	23.2	77	0	38000	Baik	tidak
12/25/2011	23.1	91	8.9	38400	Baik	tidak
1/6/2012	22.2	89	1.6	34600	Baik	tanam
1/10/2012	22.5	88	3.6	28000	Baik	tidak
1/14/2012	23.3	88	4.2	28000	Baik	tidak
1/18/2012	23.5	89	2.3	27000	Baik	tidak
1/22/2012	24.2	84	0	20600	Baik	tanam
1/26/2012	22	93	9	19600	Tidak	tidak

**Gambar 5.** Data yang di import ke rapid miner

Berikutnya proses pengolahan data menggunakan algoritma C4.5 untuk melihat pohon keputusan dan rule.



**Gambar 6.** Alur pohon keputusan C4.5

Setelah dilakukan perhitungan metode algoritma C4.5 pada *Rapid Miner*, maka terbentuk pohon keputusan seperti gambar di bawah ini:



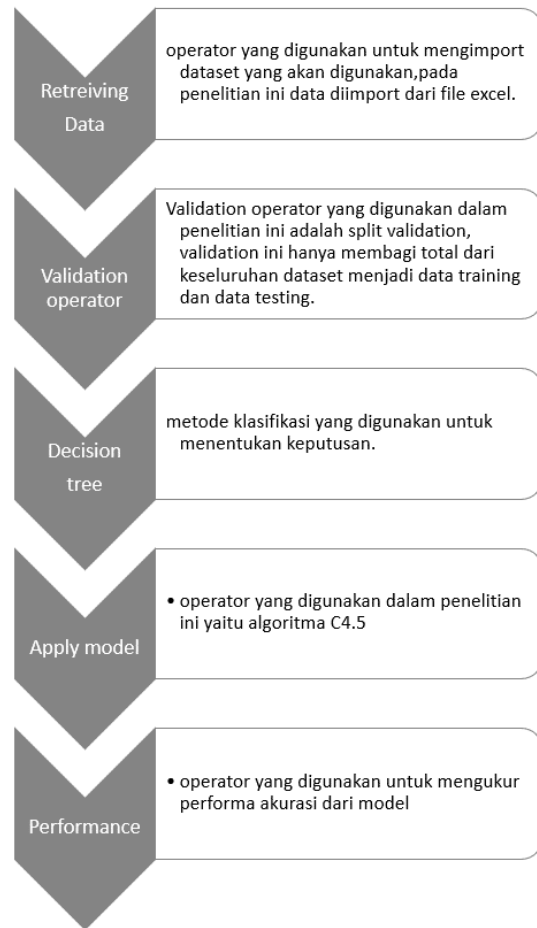
**Gambar 7.** Pohon keputusan hasil rapid miner

**Pengujian**

Pengujian ini dilakukan untuk mengetahui kinerja dari metode algoritma C4.5 dalam melakukan klasifikasi terhadap kelas yang telah ditentukan dalam uji coba ini.

Algoritma C4.5 melakukan training terhadap data-data yang telah dibagi oleh *cross validation* menjadi dua kotak yaitu *training* dan *testing*. *Training* terdiri dari *decision tree*, sedangkan *testing* terdiri dari *apply model* dan *performance*.

Berikut dijelaskan parameter dan operator yang digunakan pada model algoritma C4.5:



**Gambar 8.** Bagan model algoritma C4.5

Dalam pengujian ini akan menghasilkan nilai akurasi, *precision*, *recall* dan *AUC (Area Under Curve)*. Hasil akurasi *confusion matrix* pada pengujian sebagai berikut :

accuracy: 74.29% +/- 1.22% (micro average: 74.30%)

	true tidak	true tanam	class precision
pred tidak	393	130	75.14%
pred tanam	7	3	30.00%
class recall	99.25%	2.26%	

**Gambar 9.** Hasil akurasi *confusion matrix*

Berikut adalah perhitungan akurasi :

$$\text{Akurasi} = \frac{tp+tn}{tp+tn+fp+fn} \times 100 \%$$

(3)

$$= \frac{130+7}{130+7+3+393} \times 100 \%$$

(4)

$$= 74.2 \%$$

Hasil AUC (Area Under Curve) pada pengujian ini adalah sebagai berikut:



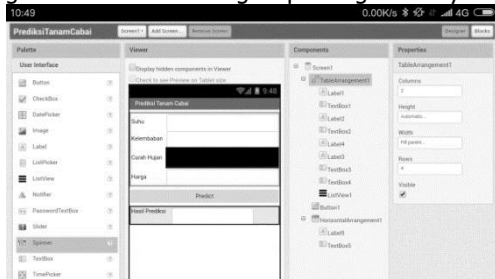
Gambar 10. Hasil AUC (Area Under Curve)

Dari Grafik pengujian AUC didapat bahwa hasil 0.947 berarti pengujian tersebut termasuk *good classification*.

**Implementasi**

Berdasarkan proses pengolahan data menggunakan Algoritma C4.5 yang telah menghasilkan pohon keputusan dan rule yang telah terbentuk, selanjutnya diimplementasikan rule tersebut dengan membuat program prediksi data pada aplikasi mobile android, dengan menggunakan tools MIT AppInventor.

Berikut ini tampilan desain tempat muka dan juga desain blok-blok logika pemrogramannya.

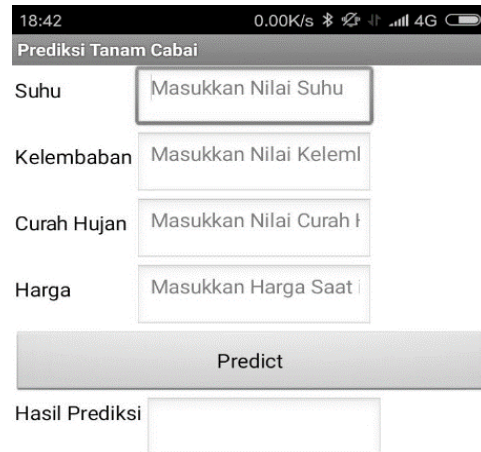


Gambar 11. Desain tampilan aplikasi

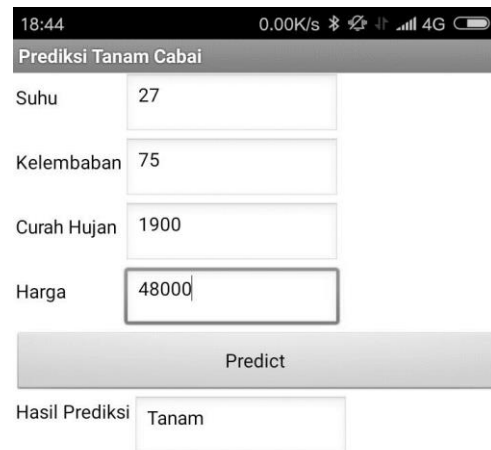


Gambar 12. Blok logika aplikasi

Hasil dari desain dan pengerjaan aplikasi mobile android tersebut adalah sebagai berikut:



Gambar 13. Aplikasi prediksi taman cabai berbasis android



Gambar 14. Prediksi aplikasi

**V. KESIMPULAN DAN SARAN**

**Kesimpulan**

Dari hasil penelitian yang telah dilakukan oleh peneliti, maka dapat disimpulkan bahwa prediksi tanam cabai dengan memanfaatkan teknik data mining menggunakan algoritma decision tree C4.5 untuk memprediksi (menentukan kelas) dengan data training yang telah diperoleh hasil akurasi 72.2 % dengan criteria akurasi *Good classification* menggunakan *confusion matrix*

**Saran**

Untuk meningkatkan kinerja dan menyempurnakan penelitian yang telah dibuat, peneliti memberikan saran sebagai berikut:

1. Penelitian ini dapat dikembangkan dengan menggabungkan atau membandingkan dengan algoritma klasifikasi lain untuk mendapatkan hasil prediksi yang lebih baik.

2. Sebaiknya jumlah data ditambah, sehingga dapat diperoleh hasil akurasi

#### **DAFTAR PUSTAKA**

- Castleman, Kenneth R., (2004), *Digital Image Processing*, Vol. 1, Ed.2, Prentice Hall, New Jersey.
- Gonzales, R., P. (2004), *Digital Image Processing (Pemrosesan Citra Digital)*, Vol. 1, Ed.2, diterjemahkan oleh Handayani, S., Andri Offset, Yogyakarta.
- Ivan, A.H., (2005), *Desain target optimal, Laporan Penelitian Hibah Bersaing*, Proyek Multitahun, Dikti, Jakarta.
- Prasetya, E., (2006), *Case Based Reasoning untuk mengidentifikasi kerusakan bangunan*, Tesis, Program Pasca Sarjana Ilmu Komputer, Univ. Gadjah Mada, Yogyakarta.
- Wyatt, J. C, dan Spiegelhalter, D., (1991), *Field Trials of Medical Decision-Aids: Potential Problems and Solutions*, Clayton, P. (ed.): *Proc. 15th Symposium on Computer Applications in Medical Care*, Vol 1, Ed. 2, McGraw Hill Inc, New York.
- Wyatt, J. C, Spiegelhalter, D, (2008), *Field Trials of Medical Decision-Aids: Potential Problems and Solutions*, Proceeding of 15th Symposium on Computer Applications in Medical Care, Washington, May 3.
- Wallace, V. P., Bamber, J. C. dan Crawford, D. C. (2000). *Classification of reflectance spectra from pigmented skin lesions, a comparison of multivariate discriminate analysis and artificial neural network*. *Journal Physical Medical Biology*, No.45, Vol.3, 2859-2871.
- Yusoff, M, Rahman, S.A., Mutalib, S., and Mohammed, A, (2006), *Diagnosing Application Development for Skin Disease Using Backpropagation Neural Network Using Backpropagation Neural Network Technique*, *Journal of Information Technology*, vol 18, hal 152-159.