

Klasifikasi Tanaman berdasarkan Fitur Bentuk dan Tekstur pada Daun menggunakan Decision Tree

Ratih Kartika Dewi
Universitas Brawijaya
ratihkartikad@ub.ac.id

Nanik Suciati
Institut Teknologi Sepuluh Nopember
nanik@if.its.ac.id

ABSTRAK

Pengenalan pola daun untuk klasifikasi jenis tanaman secara otomatis diperlukan untuk mengatasi masalah pengenalan pola daun secara manual serta mempersingkat waktu identifikasi daun. Fitur yang digunakan untuk identifikasi daun haruslah dapat merepresentasikan keadaan daun dan menjadi pembeda antar jenis tanaman. Pada penelitian ini fitur yang digunakan untuk pengenalan pola daun adalah fitur bentuk dan tekstur. Fitur bentuk dilakukan melalui fitur global, yaitu *aspect ratio*, *rectangularity*, *convex area ratio*, *circularity* sedangkan ekstraksi fitur tekstur dilakukan sesuai deskriptor lokal, yaitu *kontras*, *correlation*, *energy*, *homogeneity*, *maximum probability*, dan *entropy*. Deskriptor lokal dan fitur global tidak dapat berdiri sendiri untuk pengenalan pola daun karena 2 jenis daun yang memiliki bentuk daun yang sama dapat memiliki tekstur berbeda. Setelah proses ekstraksi fitur, metode klasifikasi dilakukan untuk mengklasifikasikan jenis daun. Pada penelitian mengenai fitur global dan deskriptor lokal, klasifikasi dilakukan menggunakan *nearest neighbor classifier*. *Nearest neighbor* memiliki kekurangan yaitu hasil klasifikasi bergantung pada jumlah *neighbor*, oleh karena itu diperlukan *classifier* yang bersifat lebih general. Dalam penelitian ini diusulkan metode *decision tree* untuk klasifikasi tanaman karena metode tersebut tidak memiliki ketergantungan dengan variabel lain. Pengujian dilakukan untuk mengetahui pengaruh variasi nilai *k* dan variasi penghitungan jarak terhadap hasil klasifikasi. Hasil dari penelitian ini adalah metode klasifikasi *nearest neighbor* menghasilkan nilai akurasi berbeda yang bergantung pada nilai *k*, dengan akurasi tertinggi pada nilai *k*=4 yaitu 47 % dan akurasi terendah pada nilai *k*=9 dan *k*=10 yaitu 40 % serta akurasi rata-rata 38,4 % sedangkan *decision tree* menghasilkan akurasi 80 %. Variasi penghitungan jarak tidak berpengaruh terhadap hasil klasifikasi.

Kata Kunci

Klasifikasi tanaman, daun, bentuk, tekstur, dan *decision tree*.

1. PENDAHULUAN

Daun adalah bagian tanaman yang bisa digunakan untuk klasifikasi spesies tanaman. Masing-masing daun membawa informasi substansial yang dapat digunakan untuk identifikasi dan klasifikasi beragam jenis tanaman. Pengenalan pola daun secara otomatis diperlukan untuk mempersingkat waktu identifikasi daun dan masalah lain yang timbul saat menggunakan pengenalan pola daun secara manual melalui botanis.

Proses pengenalan pola secara normal terjadi pada tahap *preprocessing* yang kemudian diikuti dengan proses ekstraksi fitur. Setelah proses tersebut barulah klasifikasi dilakukan dengan proses pembelajaran dari basis data fitur daun yang ada. Untuk membentuk sistem pengenalan pola daun, fitur yang digunakan haruslah dapat merepresentasikan keadaan daun dan dapat menjadi pembeda antar jenis tanaman.

Daun memiliki fitur yang berbeda antar satu jenis dengan jenis lainnya. Fitur tersebut membawa informasi signifikan yang dapat membantu manusia mengenali dan mengklasifikasikan

tanaman dengan melihatnya. Perkembangan penelitian mengenai klasifikasi jenis tanaman berkembang pada fitur yang digunakan dan metode klasifikasi yang dipilih. Fitur bentuk pada daun adalah fitur yang signifikan yang banyak digunakan manusia (Ab Jabal, 2013). Penelitian (Wu et al, 2007) menyatakan bahwa fitur bentuk seperti *diameter*, *physiological length*, *physiological width*, *area* dan *perimeter* adalah informasi geometri dasar yang dapat diekstraksi dari daun. Metode ekstraksi fitur tekstur lain adalah melalui *moment invariants*, yang pernah digunakan dalam penelitian (Zulkifli et al., 2011) menggunakan 10 kelas daun. Sedangkan penelitian (Shabanzade, 2011) menyatakan bahwa secara umum fitur bentuk yang dapat diekstraksi dari daun terdiri dari panjang dan lebar dari luas daun atau sifat-sifat lain yang menentukan bentuk secara universal.

Fitur bentuk tidak dapat berdiri sendiri untuk pengenalan pola daun karena 2 jenis daun memiliki bentuk daun yang sama tetapi dapat memiliki tekstur berbeda (Shabanzade, 2011). Berikut adalah alasan penggunaan paduan fitur global dan deskriptor lokal secara bersamaan (Shabanzade, 2011):

- Citra daun memiliki properti yang berbeda karena kondisi berbeda, misalnya warna kuning saat musim gugur. Kondisi tersebut menyebabkan penghitungan yang salah pada deskriptor lokal.
- Bentuk daun dapat tidak penuh, misalnya jika dimakan tanaman. Kondisi seperti ini membuat perhitungan dengan fitur global menghasilkan hasil yang salah.
- Beberapa daun dapat mengalami sakit yang berupa lubang atau titik pada daun. Kondisi ini menyebabkan deskriptor lokal menghasilkan nilai yang tidak representatif dan menyebabkan kesalahan dalam pengenalan pola.

Oleh karena itu perlu penambahan fitur tekstur. Fitur tekstur untuk pengenalan pola daun yang bisa digunakan adalah *Gray-level occurrence matrices*, *Gabor Filter*, and *Local binary pattern* (Kadir et al., 2011). Penelitian (Shabanzade, 2011) menggunakan ekstraksi fitur tekstur meliputi *kontras*, *correlation*, *energy*, *homogeneity*, *maximum probability*, dan *entropy* dan menggabungkannya dengan ekstraksi fitur bentuk melalui *aspect ratio*, *rectangularity*, *convex area ratio*, *circularity*.

Pada penelitian (Shabanzade, 2011), pengenalan pola daun untuk klasifikasi tanaman menggunakan ekstraksi fitur bentuk dan tekstur seperti yang telah dideskripsikan sebelumnya, kemudian metode *LDA (Linear Discriminant Analysis)* digunakan untuk menyeleksi fitur yang paling berpengaruh terhadap hasil klasifikasi. Metode klasifikasi dilakukan menggunakan metode *nearest neighbor*. Metode klasifikasi *k-nearest neighbor* memiliki kekurangan yaitu hasil klasifikasi bergantung pada jumlah *k*, oleh karena itu diperlukan metode klasifikasi yang bersifat lebih umum (Luukka, 2009).

Pada penelitian ini fitur yang digunakan untuk pengenalan pola daun adalah fitur bentuk dan tekstur. Fitur bentuk dilakukan melalui fitur global, yaitu *aspect ratio*, *rectangularity*, *convex*

area ratio, *circularity* sedangkan ekstraksi fitur tekstur dilakukan sesuai deskriptor lokal, yaitu kontras, *correlation*, *energy*, *homogeneity*, *maximum probability*, dan *entropy*. Setelah ekstraksi fitur dilakukan, maka selanjutnya adalah melakukan klasifikasi.

Dalam penelitian ini diusulkan metode klasifikasi *decision tree* untuk klasifikasi tanaman karena metode tersebut tidak memiliki ketergantungan dengan variabel lain. Pengujian sistem dalam penelitian ini dilakukan dengan menguji pada membandingkan metode klasifikasi *decision tree* dan *k-nearest neighbor* dengan variasi nilai *k* mulai dari 1 sampai 10 dan variasi perhitungan jarak, melalui perhitungan jarak Euclidean, Cityblock, Cosine, Correlation, dan Hamming. Penelitian klasifikasi tanaman menggunakan fitur bentuk dan tekstur dengan fitur bentuk dan tekstur menggunakan *decision tree* diharapkan dapat mengklasifikasi citra daun yang dimasukkan ke dalam sistem secara benar dan metode klasifikasi tidak memiliki ketergantungan terhadap variabel lain.

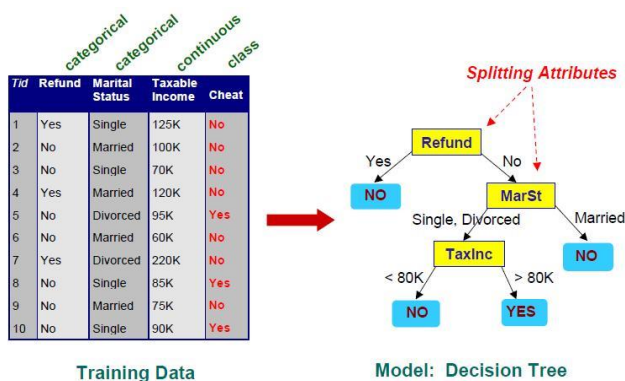
Penelitian ini memiliki struktur bagian sebagai berikut, yaitu bagian I pendahuluan yang berisi latar belakang penelitian, bagian II berisi studi literatur, bab III arsitektur sistem, bagian IV berisi analisis dan pengujian system, bagian V berisi kesimpulan, bagian VI berisi saran dan bagian terakhir merupakan pustaka yang dipakai sebagai acuan dalam penelitian ini.

2. STUDI LITERATUR

2.1 Decision Tree

Konsep dasar klasifikasi yang dilakukan dengan decision tree selalu diawali dengan pengajuan pertanyaan untuk penentuan kelompok. Dalam decision tree, pertanyaan pertama akan menempati level 0 atau simpul akar. Pertanyaan berikutnya pada level 1 atau cabang simpul akar. Dalam setiap simpul yang ditanyakan adalah nilai suatu atribut. Jawaban tiap pertanyaan diungkapkan dalam cabang-cabang simpul. Pertanyaan akan berakhir jika sudah didapatkan dengan jelas kelas atau obyek yang dicari. Metode untuk mengukur kesamaan node menggunakan konsep *information gain* atau *entropy reduction* untuk memilih split yang optimal (Tan, 2006).

Gambar 1 menunjukkan decision tree untuk menentukan keputusan apakah seseorang layak untuk mendapatkan pinjaman dari bank dengan variable masukan berupa *refund* (apakah seseorang yang pernah meminjam uang sudah membayar uang yang dipinjamnya), status pernikahan (*marital status*) dan penghasilan kena pajak (*taxable income*).



Gambar 1 Klasifikasi dengan Decision Tree (Tan, 2006)

Cara kerja decision tree membentuk tree dengan split atribut berdasarkan information gain. Dalam Gambar 1 kita memiliki 3

atribut, yaitu *refund*, *marital status* dan *taxable income* serta 1 kelas, yaitu *cheat* yang menyatakan potensi kecurangan dari seseorang berdasarkan kombinasi nilai 3 atribut tersebut. Untuk memilih atribut yang akan displit, mulanya kita menghitung entropy dengan Persamaan 2.1, dimana variable X, memiliki k nilai kemungkinan dengan probabilitas p_1, p_2, \dots, p_k .

$$H(X) = -\sum_j p_j \log_2(p_j) \quad (1)$$

Pada Gambar 7, ada 7 kelas bernilai no dan 3 kelas bernilai yes dari jumlah keseluruhan data yang berjumlah 10. Entropy dapat dihitung seperti Persamaan 2.2

$$H(T) = -\frac{7}{10} \log\left(\frac{7}{10}\right) - \frac{3}{10} \log\left(\frac{3}{10}\right) \quad (2)$$

Misalkan kita memiliki suatu kandidat split S, yang mempartisi data training T ke beberapa subset, T1, T2,...Tk. Syarat informasi yang dipentingkan dapat dihitung:

$$H_S(T) = \sum_{i=1}^k P_i H_S(T_i) \quad (3)$$

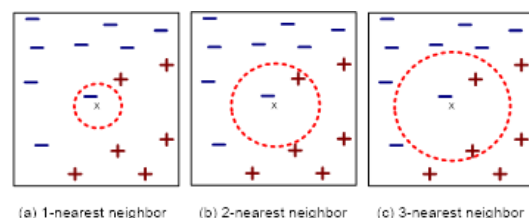
Dimana P_i merepresentasikan proporsi dari record dalam subset i. Sehingga information gain dapat didefinisikan:

$$\text{gain}(S) = H(T) - H_S(T) \quad (4)$$

Atribut dengan nilai information gain yang paling tinggi dipilih menjadi splitting atribut dari akar seperti Gambar 7. Setelah terbentuk decision tree, kita bisa membuat decision rules, misalnya Jika *refund=no* dan *marital status=married*, maka potensi cheat dari orang tersebut tidak ada sehingga dia bisa menjadi kandidat yang layak untuk mendapatkan pinjaman dari bank.

2.2 K-nearest neighbor

Konsep dasar dari algoritma K-nearest neighbor (KNN) adalah menentukan kelas berdasarkan tingkat kesamaan yang tinggi dari data latih. Langkah pertama penghitungan KNN yaitu menghitung jarak ke data latih lain, mengidentifikasi k-tetangga terdekat dan menggunakan kelas label dari tetangga terdekat untuk menentukan label kelas dari record yang belum diketahui, misalnya dengan mengambil suara terbanyak (Tan, 2006). K- Nearest neighbor dari suatu record x merupakan titik-titik data yang memiliki k jarak terkecil ke x, dimana x adalah data yang dicari kelasnya. Pada Gambar 2, terlihat variasi nilai k dan ilustrasi KNN dalam gambar.



Gambar 2 K-nearest neighbor

Fungsi jarak yang paling umum digunakan adalah jarak Euclidean, tetapi tidak menutup kemungkinan perhitungan jarak dengan metode lainnya dapat dipakai. Metode pengukuran jarak lainnya misalnya cityblock, cosine, correlation dan hamming.

$$d_{Euclidean}(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (5)$$

$$d_{cityblock}(x, y) = \sum_i |x_i - y_i| \quad (6)$$

$$d_{cosine}(x, y) = \frac{\sum x \cdot y}{\sqrt{\sum x^2} \sqrt{\sum y^2}} \quad (7)$$

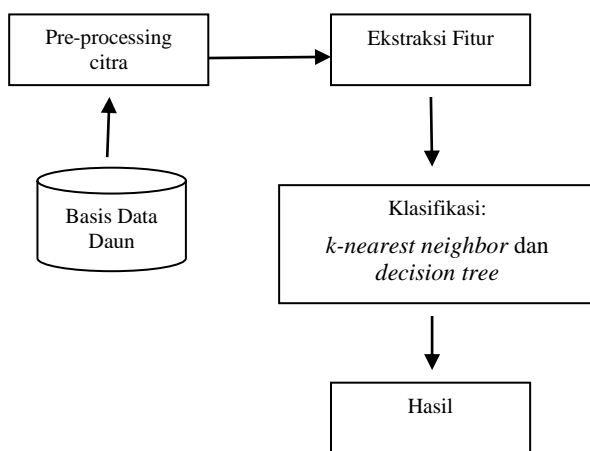
$$d_{correlation}(x, y) = \frac{dCov(x, y)}{\sqrt{dVar(x) dVar(y)}} \quad (8)$$

$$d_{hamming}(x, y) = \sum |x - y| \quad (9)$$

Dimana, $x=x_1, x_2, \dots, x_m$, dan y_1, y_2, \dots, y_m merepresentasikan nilai atribut m dari dua *rechord*.

3. ARSITEKTUR SISTEM

Penelitian ini memiliki arsitektur sistem seperti ditunjukkan pada Gambar 3.



Gambar 3 Arsitektur Sistem

Pada tahap pertama, *pre-processing* yang dilakukan pada citra daun berfungsi untuk menyiapkan citra agar bisa diproses ke tahap selanjutnya, yaitu ekstraksi fitur. Pada tahap ekstraksi fitur, fitur daun yang diekstraksi berupa fitur bentuk dan fitur tekstur. Setelah melalui tahap ekstraksi fitur dan mendapatkan fitur yang diinginkan, tahap selanjutnya adalah tahap klasifikasi yang berfungsi untuk memilah beragam daun ke dalam kelas yang cocok, dalam hal ini adalah spesies tanaman. Pada tahap klasifikasi, *classifier* yang digunakan adalah *k-nearest neighbor* dan *decision tree*. Setelah memperoleh hasil dari proses klasifikasi, maka dilakukan pengujian melalui pengujian akurasi beserta analisisnya.

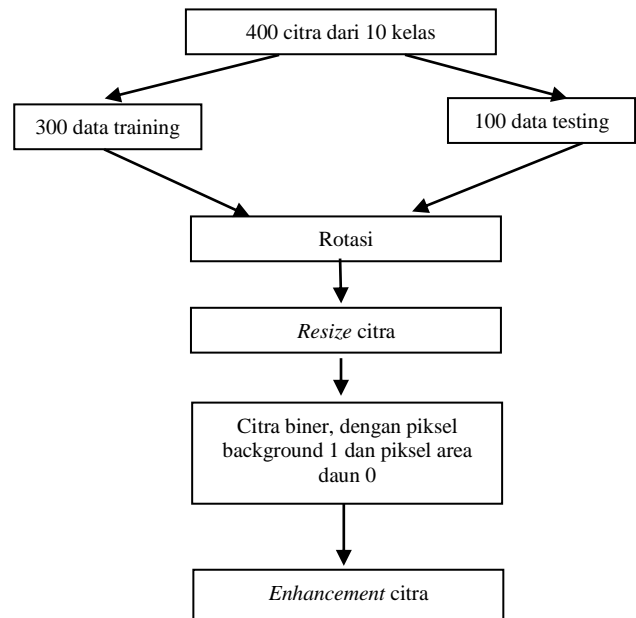
3.1 Pre-processing

Dalam tahap *pre-processing*, ada beberapa langkah yang dilakukan untuk kemudahan pemrosesan data pada tahap selanjutnya, yaitu proses ekstraksi fitur. Dalam bentuk ringkas, proses *pre-processing* citra pada penelitian ini digambarkan pada Gambar 4.

Data citra daun yang akan digunakan didapatkan dari dataset *flavia*. Citra daun yang digunakan dalam penelitian ini berjumlah 400 dengan 10 macam kelas tanaman, data ini di-*split* menjadi 2 bagian yaitu 300 untuk data training dan 100 untuk data testing. Sebanyak 300 citra dari 10 kelas digunakan

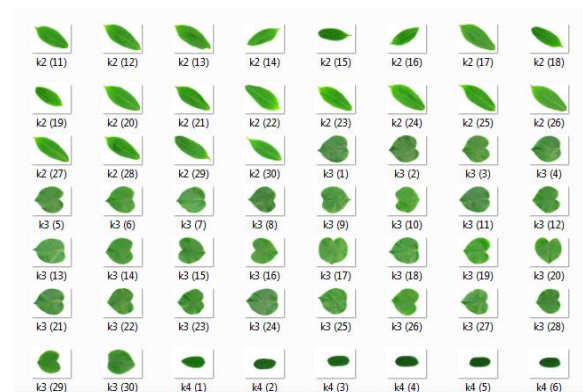
untuk data training dan 100 citra dari 10 kelas digunakan untuk data testing. Dataset dapat diunduh pada (Baosheng, 2015).

Seluruh data kemudian dilakukan rotasi untuk penyeragaman posisi. Setelah proses rotasi, pengukuran ulang dilakukan terhadap citra sehingga diperoleh citra dengan ukuran yang lebih kecil dari citra awal. Tahap selanjutnya adalah membuat citra biner dengan *thresholding* Otsu. Tahap terakhir adalah tahap *enhancement* dengan menyeimbangkan kontras citra.



Gambar 4. Pre-processing Citra

Gambar 5 adalah contoh data daun yang digunakan dalam penelitian ini, meliputi data training dan data testing. Citra tersebut kemudian dirotasi agar memiliki penampakan yang sama.

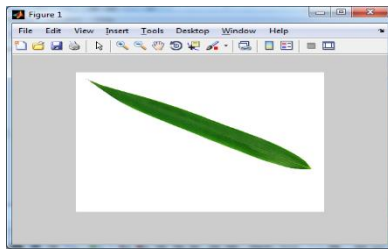


Gambar 5. Contoh Data Training dan Testing

Gambar 6 merupakan citra hasil rotasi. Setelah dilakukan rotasi, ukuran citra diperkecil karena file berukuran besar. Gambar diperkecil sampai dengan 20% dan metode yang digunakan adalah metode *nearest* pada fungsi *imresize* di MATLAB.



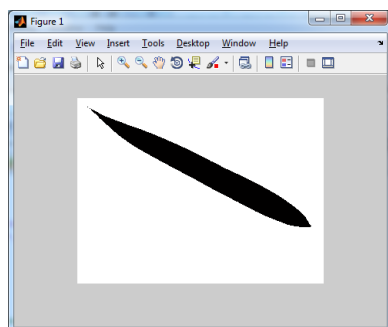
Gambar 6. Hasil dari Split Data Setelah Rotasi



Gambar 7. Citra hasil imresize

Gambar 7 menunjukkan salah satu citra yang ukurannya telah diperkecil. Untuk proses ekstraksi fitur, citra harus dalam bentuk biner dengan piksel background 1 (putih) dan piksel area daun 0 (hitam). Metode yang digunakan adalah *thresholding* adaptif menggunakan *thresholding* Otsu. *Thresholding* Otsu menggunakan metode *clustering* untuk proses threshold. Algoritma ini mengasumsikan citra yang akan di-threshold memiliki 2 kelas piksel (*bimodal histogram-foreground* dan *bimodal histogram-background*) kemudian menghitung threshold optimal yang memisahkan kedua kelas tersebut.

Citra hasil binerisasi kemudian diberikan teknik *image enhancement* dengan memperbaiki kontras citra. *Imadjust* adalah fungsi MATLAB yang digunakan untuk meningkatkan kontras dari citra dengan memetakan nilai intensitas citra ke dalam nilai baru sehingga 1% dari data tersaturasi pada intensitas rendah dan tinggi. Gambar 8 menunjukkan citra hasil binerisasi. Objek daun diberikan nilai piksel 0 sehingga berwarna hitam dan *background* diberikan nilai 1 sehingga berwarna putih.



Gambar 8. Citra hasil binerisasi

3.2 Ekstraksi Fitur

Ekstraksi fitur pada daun dilakukan menggunakan fitur bentuk dan tekstur. Ekstraksi fitur bentuk dilakukan sesuai fitur global, yaitu *aspect ratio*, *rectangularity*, *convex area ratio*, *circularity* sedangkan ekstraksi fitur tekstur dilakukan sesuai deskriptor

lokal, yaitu kontras, *correlation*, *energy*, *homogeneity*, *maximum probability*, dan *entropy*.

3.2.1 Tekstur

Deskriptor lokal digunakan untuk merepresentasikan tekstur daun. Deskriptor lokal yang digunakan untuk ekstraksi fitur meliputi kontras, *correlation*, *energy*, *homogeneity*, *maximum probability*, dan *entropy* (Shabanzade, 2011). Jika G_{ij} adalah jumlah berapa kali piksel dengan tingkat keabuan i terjadi berdampingan dengan tingkat keabuan j dan n adalah jumlah pasangan piksel maka P_{ij} yang merupakan distribusi probabilitas bersama dari pasangan piksel dengan tingkat keabuan i dan tingkat keabuan j dapat dihitung dengan membagi G_{ij} dengan n . Penghitungan piksel yang berdampingan memiliki arah dan jarak, arah sesuai sudut 0° , 45° , 90° , dan 135° dan jarak adalah jarak antar piksel, misalnya 1 atau 2. Penghitungan yang digunakan dalam penelitian ini menggunakan P_{ij} sebagai dasar penghitungan dapat dijelaskan sebagai berikut:

1. Kontras

Kontras dari sebuah piksel dan tetangganya yang dihitung untuk semua piksel pada citra dapat dihitung dengan Persamaan 10 dimana P_{ij} merupakan distribusi probabilitas bersama dari pasangan piksel dengan tingkat keabuan i dan tingkat keabuan j .

$$f1 = \sum_{i=1}^k \sum_{j=1}^k (i - j)^2 P_{ij} \quad (10)$$

2. Correlation

Correlation adalah pengukuran korelasi antara satu piksel dengan tetangganya yang dihitung untuk semua piksel pada citra. Jika m adalah mean (nilai rata-rata) piksel pada posisi x dan y dan σ adalah standar deviasi, maka *correlation* dapat dihitung berdasarkan Persamaan 11 sebagai berikut:

$$f2 = \sum_{i=1}^k \sum_{j=1}^k \frac{(i-mx)(j-my)P_{ij}}{\sigma_x \sigma_y} \quad (11)$$

3. Energy

Energy adalah penjumlahan dari elemen *Co-occurrence matrix* yang dihitung berdasarkan Persamaan 12 dimana P_{ij} merupakan distribusi probabilitas bersama dari pasangan piksel dengan tingkat keabuan i dan tingkat keabuan j .

$$f3 = \sum_{i=1}^k \sum_{j=1}^k P_{ij}^2 \quad (12)$$

4. Homogeneity

Homogeneity adalah kesamaan (*similarity*) dari *co-occurrence matrix* dan *diagonal matrix* yang dihitung berdasarkan Persamaan 13 dimana P_{ij} merupakan distribusi probabilitas bersama dari pasangan piksel dengan tingkat keabuan i dan tingkat keabuan j .

$$f4 = \sum_{i=1}^k \sum_{j=1}^k \frac{P_{ij}}{1+|i-j|} \quad (13)$$

5. Maximum probability

Maximum probability adalah indikasi respons paling kuat terhadap operasi yang dilakukan terhadap citra. Maximum probability dapat dihitung berdasarkan Persamaan 14:

$$f5 = \text{Max}_{i,j} (P_{ij}) \quad (14)$$

6. Entropy

Entropy adalah rasio randomisasi pada citra yang dapat dihitung berdasarkan Persamaan 15 berikut:

$$f6 = - \sum_{i=1}^k \sum_{j=1}^k P_{ij} \log_2(P_{ij}) \quad (15)$$

3.2.2 Bentuk

Fitur global digunakan untuk merepresentasikan bentuk daun. Fitur global yang digunakan untuk ekstraksi fitur meliputi *aspect ratio*, *rectangularity*, *convex area ratio*, *circularity* (Shabanzade, 2011):

1. Aspect ratio

Aspect ratio adalah rasio antara L_{\max} (panjang maksimum) dan L_{\min} (panjang minimum) pada minimum *bounding elips* pada citra daun yang dirumuskan pada Persamaan 16.

$$f7 = \frac{L_{\max}}{L_{\min}} \quad (16)$$

2. Rectangularity

Rectangularity adalah rasio antara *region of interest* (ROI) dalam hal ini daun dengan *aspect ratio* yang dirumuskan pada Persamaan 17.

$$f8 = \frac{A_{ROI}}{f7} \quad (17)$$

3. Convex area ratio

Convex area ratio adalah rasio antara area ROI dan *convex hull area* (A_c) yang dirumuskan pada Persamaan 18.

$$f9 = \frac{A_c}{A_{ROI}} \quad (18)$$

4. Circularity

Circularity adalah pembagian antara mean ROI dan variance ROI yang dirumuskan pada Persamaan 19.

$$f10 = \frac{\mu_{ROI}}{\sigma_{ROI}} \quad (19)$$

3.3 Klasifikasi dengan Decision Tree

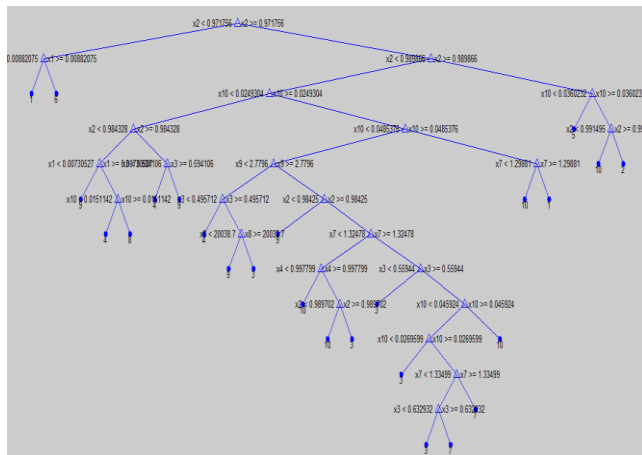
Berdasarkan konsep pembentukan decision tree pada bagian 2.1, kita dapatkan decision tree untuk klasifikasi tanaman

seperti terlihat pada Gambar 9 sehingga terbentuk decision rules sebagai berikut:

- Jika $x_2 < 0,971756$ dan $x_1 \leq 0,00882075$ maka kelas=1
- Jika $x_2 < 0,971756$ dan $x_1 > 0,00882075$ maka kelas=6
- Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} < 0,0249304$, $x_2 < 0,984328$, $x_1 < 0,00730527$ maka kelas=9
- Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} < 0,0249304$, $x_2 < 0,984328$, $x_1 < 0,00730527$ dan $x_{10} < 0,0151142$ maka kelas=4
- Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} < 0,0249304$, $x_2 < 0,984328$, $x_1 < 0,00730527$ dan $x_{10} \geq 0,0151142$ maka kelas=8
- Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} < 0,0249304$, $x_2 < 0,984328$, $x_3 < 0,594106$ maka kelas=4.
- Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} < 0,0249304$, $x_2 < 0,984328$, $x_3 \geq 0,594106$ maka kelas=9.
- Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} \geq 0,0249304$, $x_{10} < 0,0485376$, $x_9 < 2,7796$, $x_3 < 0,495712$ maka kelas=4
- Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} \geq 0,0249304$, $x_{10} < 0,0485376$, $x_9 < 2,7796$, $x_3 < 0,495712$ dan $x_8 < 20038,7$ maka kelas=9
- Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} \geq 0,0249304$, $x_{10} < 0,0485376$, $x_9 < 2,7796$, $x_3 < 0,495712$ dan $x_8 \geq 20038,7$ maka kelas=3
- Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} \geq 0,0249304$, $x_{10} < 0,0485376$, $x_9 < 2,7796$, dan $x_2 < 0,98425$ maka kelas=9
- Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} \geq 0,0249304$, $x_{10} < 0,0485376$, $x_9 < 2,7796$, dan $x_2 \geq 0,98425$, $x_7 < 1,32478$, $x_4 < 0,997799$ maka kelas=10
- Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} \geq 0,0249304$, $x_{10} < 0,0485376$, $x_9 < 2,7796$, dan $x_2 \geq 0,98425$, $x_7 < 1,32478$, $x_4 \geq 0,997799$ dan $x_2 \geq 0,989702$ maka kelas=3
- Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} \geq 0,0249304$, $x_{10} < 0,0485376$, $x_9 < 2,7796$, dan $x_2 \geq 0,98425$, $x_7 < 1,32478$, $x_4 \geq 0,997799$ dan $x_2 < 0,989702$ maka kelas=10
- Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} \geq 0,0249304$, $x_{10} < 0,0485376$, $x_9 < 2,7796$, dan $x_2 \geq 0,98425$, $x_7 < 1,32478$, $x_3 < 0,55944$ maka kelas=3
- Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} \geq 0,0249304$, $x_{10} < 0,0485376$, $x_9 < 2,7796$, dan $x_2 \geq 0,98425$, $x_7 < 1,32478$, $x_3 \geq 0,55944$, $x_{10} < 0,045924$ dan $x_{10} < 0,0269599$ maka kelas=3
- Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} \geq 0,0249304$, $x_{10} < 0,0485376$, $x_9 < 2,7796$, dan $x_2 \geq 0,98425$, $x_7 < 1,32478$, $x_3 \geq 0,55944$, $x_{10} < 0,045924$, $x_{10} < 0,0269599$, $x_7 < 1,33499$, $x_3 < 0,632932$ maka kelas=3
- Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} \geq 0,0249304$, $x_{10} < 0,0485376$, $x_9 < 2,7796$, dan $x_2 \geq 0,98425$, $x_7 < 1,32478$, $x_3 \geq 0,55944$, $x_{10} < 0,045924$, $x_{10} < 0,0269599$, $x_7 \geq 1,33499$ maka kelas=7

- s) Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} \geq 0,0249304$, $x_{10} < 0,0485376$, $x_9 < 2,7796$, dan $x_2 \geq 0,98425$, $x_7 < 1,32478$, $x_3 \geq 0,55944$, $x_{10} < 0,045924$, $x_{10} < 0,0269599$, $x_7 < 1,33499$, $x_3 \geq 0,632932$ maka kelas=7
- t) Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} \geq 0,0249304$, $x_{10} < 0,0485376$, $x_7 < 1,29881$ maka kelas=10
- u) Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} \geq 0,0249304$, $x_{10} < 0,0485376$, $x_7 \geq 1,29881$ maka kelas=1
- v) Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} < 0,0360232$ maka kelas=5
- w) Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} < 0,0360232$ dan $x_2 < 0,991$ maka kelas=10
- x) Jika $x_2 \geq 0,971756$, $x_2 < 0,989866$, $x_{10} < 0,0360232$ dan $x_2 < 0,991$ maka kelas=2

Dimana x adalah fitur bentuk atau tekstur dengan x1 adalah kontras, x2 correlation, x3 energy, x4 homogeneity, x5 maximum probability, x6 entropy, x7 aspect ratio, x8 rectangularity, x9 convex area ratio dan x10 adalah circularity.



Gambar 9. Decision Tree Klasifikasi Tanaman

Apabila ada data baru yang tidak diketahui kelasnya (tanaman dengan spesies apa), maka kita bisa memprediksi spesies tersebut berdasarkan decision tree yang telah terbentuk seperti pada Gambar 9.

3.4 Klasifikasi dengan K nearest neighbor

Klasifikasi dengan KNN sesuai konsep klasifikasi pada bagian 2.2, dalam penelitian ini digunakan nilai k mulai dari 1 sampai dengan 10 dan jarak Euclidean, cityblock, cosine, correlation dan hamming.

4. ANALISIS DAN PENGUJIAN

Citra daun dengan 10 macam kelas digunakan untuk data training sebanyak 300 data dan 100 data untuk data testing. Tahap *pre-processing* memperbaiki kualitas citra dan menghasilkan citra biner yang kemudian diproses ke dalam tahap selanjutnya, yaitu ekstraksi fitur.

Pada penelitian ini fitur yang digunakan untuk pengenalan pola daun adalah fitur bentuk dan tekstur. Fitur bentuk dilakukan melalui fitur global, yaitu *aspect ratio*, *rectangularity*, *convex area ratio*, *circularity* sedangkan ekstraksi fitur tekstur dilakukan sesuai deskriptor lokal, yaitu kontras, *correlation*, *energy*, *homogeneity*, *maximum probability*, dan *entropy*. Fitur warna dalam hal ini kurang representatif untuk klasifikasi

tanaman melalui daun karena warna daun relatif sama, yaitu hijau. Fitur warna lebih cocok digunakan untuk identifikasi penyakit daun karena terdapat perbedaan warna antara daun sehat dan sakit.

Setelah ekstraksi fitur dilakukan, kemudian dilakukan pengujian dengan metode klasifikasi yang berbeda, yaitu dengan membandingkan metode klasifikasi *k-nearest neighbor* dan *decision tree*. Pengujian sistem dalam penelitian ini dilakukan dengan menguji pada membandingkan metode klasifikasi *decision tree* dan *k-nearest neighbor* dengan variasi nilai k mulai dari 1 sampai 10 dan variasi perhitungan jarak, melalui perhitungan jarak Euclidean, Cityblock, Cosine, Correlation, dan Hamming. Implementasi program dilakukan dengan MATLAB dan dilakukan pengujian akurasi melalui metode klasifikasi *k-nearest neighbor* dengan variasi jarak dan nilai k dan *decision tree*. Hasil pengujian dapat dideskripsikan pada Tabel 1 sebagai berikut:

Tabel 1. Pengujian akurasi nearest neighbor

	K=1	K=2	K=3	K=4	K=5
Euclidean	42	42	44	47	43
Cityblock	42	42	44	47	43
Cosine	42	42	44	47	43
Correlation	42	42	44	47	43
Hamming	42	42	44	47	43
	K=6	K=7	K=8	K=9	K=10
Euclidean	45	42	41	40	40
Cityblock	45	42	41	40	40
Cosine	45	42	41	40	40
Correlation	45	42	41	40	40
Hamming	45	42	41	40	40

Pengujian akurasi dengan variasi nilai k dan metode penghitungan distance pada Tabel 1 menunjukkan bahwa hasil akurasi sangat bergantung pada penentuan nilai k. Nilai akurasi rata-rata untuk nilai k=1 sampai dengan 10 adalah 38,5 %. Dengan akurasi tertinggi pada nilai k=4 yaitu 47 % dan akurasi terendah pada nilai k=9 dan k=10 yaitu 40 %. Metode penghitungan jarak yang digunakan menunjukkan bahwa tidak ada pengaruh antara perbedaan metode penghitungan dengan nilai pengujian akurasi.

Dalam Tabel 2 terlihat bahwa rata-rata akurasi pada metode k-nearest neighbor yaitu 38,5 % sedangkan decision tree memberikan hasil akurasi yang lebih tinggi daripada metode k-nearest neighbor, yaitu 80% sehingga dapat menjadi solusi yang baik untuk permasalahan ini selain mengatasi permasalahan ketergantungan variabel k pada metode k-nearest neighbor.

Tabel 2. Pengujian akurasi beragam metode klasifikasi

Metode klasifikasi	Akurasi
k-Nearest neighbor	38,5 %
Decision tree	80 %

5. KESIMPULAN

Klasifikasi jenis tanaman dapat diidentifikasi melalui daun. Fitur dari daun yang dapat diekstraksi adalah fitur bentuk dan tekstur. Dalam penelitian ini diusulkan metode decision tree

untuk klasifikasi tanaman karena metode tersebut tidak memiliki ketergantungan dengan variabel lain, seperti ketergantungan nilai k pada metode klasifikasi *nearest neighbor*. Pada pengujian citra daun dengan nilai $k=1$ sampai 10 untuk metode klasifikasi *nearest neighbor*, akurasi rata-rata untuk nilai $k=1$ sampai dengan 10 adalah 38,5 %. Dengan akurasi tertinggi pada nilai $k=4$ yaitu 47 % dan akurasi terendah pada nilai $k=9$ dan $k=10$ yaitu 40 %. Metode penghitungan jarak yang digunakan menunjukkan bahwa tidak ada pengaruh antara perbedaan metode penghitungan dengan nilai pengujian akurasi.

Data daun yang diklasifikasikan melalui ekstraksi fitur bentuk dan tekstur menghasilkan akurasi 80% saat menggunakan metode klasifikasi *decision tree*, sedangkan metode klasifikasi *nearest neighbor* menghasilkan akurasi rata-rata 38,5 %. Oleh karena itu *decision tree* dapat menjadi solusi yang baik untuk mengatasi permasalahan ketergantungan variabel pada metode *k-nearest neighbor*.

6. SARAN

Penelitian lebih lanjut dapat dilakukan dengan menggunakan model klasifikasi yang berbeda untuk meningkatkan akurasi. Pada penelitian saat ini menggunakan *decision tree* didapatkan tingkat akurasi sebesar 80%, kesalahan prediksi kelas tanaman salah satunya disebabkan adanya nilai fitur yang memiliki nilai tidak jauh berbeda sehingga menghasilkan perbedaan yang tidak begitu signifikan antar kelas yang akan diprediksi. Oleh karena itu, perlu adanya variasi model klasifikasi (*classifier*) untuk meningkatkan akurasi.

7. REFERENSI

Ab Jabal, Mohamad Faizal, et al. "LEAF FEATURES EXTRACTION AND RECOGNITION APPROACHES TO

CLASSIFY PLANT." *Journal of Computer Science* 9.10 (2013): 1295.

Baosheng. [http://sourceforge.net/projects/flavia/files/Leaf%20Image %20Dataset/1.0/Leaves.tar.bz2/download](http://sourceforge.net/projects/flavia/files/Leaf%20Image%20Dataset/1.0/Leaves.tar.bz2/download). Tanggal akses: 9 Desember 2015.

Kadir, A., L.E. Nugroho, A. Susanto and P.I. Santosa, 2011. A comparative experiment of several shape methods in recognizing plants. *Int. J. Comput. Sci. Inform. Technol.*, 3: 256-263. DOI: 10.5121/ijcsit.2011.3318.

Kadir, Abdul, et al. "Leaf classification using shape, color, and texture features." *arXiv preprint arXiv:1401.4447* (2013).

Luukka, Pasi. "Similarity classifier using similarities based on modified probabilistic equivalence relations." *Knowledge-Based Systems* 22.1 (2009): 57-62.

Shabanzade, Maliheh, Morteza Zahedi, and Seyyed Amin Aghvami. "Combination of local descriptors and global features for leaf recognition." *Signal and Image Processing: An International Journal (SIPIJ)*. v2 i3 (2011): 23-31.

Wu, S.G., F.S. Bao, E.Y. Xu, Y.X. Wang and Y.F. Chang *et al.*, 2007. A leaf recognition algorithm for plant classification using probabilistic neural network. *Proceedings of the IEEE 7th International Symposium on Signal Processing and Information Technology*, Dec. 15-18, IEEE Xplore Press, Giza, pp: 11-16. DOI: 10.1109/ISSPIT.2007.4458016

Zulkifli, Z., S. Puteh and I.A. Mohtar, 2011. Plant leaf identification using moment invariants and general regression neural network. *Proceedings of the 11th International Conference on Hybrid Intelligent Systems*, Dec. 5-8, Melacca, pp: 430-435. DOI: 10.1109/HIS.2011.6122144.