

Klasifikasi Citra Menggunakan Metode Random Forest dan Sequential Minimal Optimization (SMO)

Saprudin^{a1}, Resti Amalia^{a2}, Perani Rosyani^{a3}

*^aTeknik Informatika Fakultas Teknik, Universitas Pamulang
Jl. Surya Kencana no. 1, Pamulang Barat*

¹dosen00845@unpam.ac.id

²dosen00850@unpam.ac.id

³dosen00837@unpam.ac.id

Abstrak

Klasifikasi bertujuan untuk memisahkan antara jenis-jenis kategori dan label yang telah dibuat. Didalam penelitian ini kami mencoba mengklasifikasi citra bunga dengan menggunakan proses segmentasi. Segmentasi ini kami gunakan untuk menghilangkan noise yang ada didalam background. Selain itu untuk mendapatkan fitur dari gambar yang kami klasifikasi. Fitur yang kami gunakan adalah eccentricity, perimeter, metric dan area. Setelah itu kami proses sehingga mendapatkan nilai. Kami menggunakan 120 gambar dari dataset 17flower dengan 2 jenis bunga berwarna putih dan kuning. Kemudian kami bagi menjadi data training dan data test dengan metode klasifikasi Random forest dan SMO yang diolah dengan skenario 10-fold cross-validation dan 66% split. Hasilnya adalah metode Random forest memiliki akurasi yang terbaik dengan 2 skenario yang berbeda dibandingkan dengan SMO.

Kata kunci: Random Forest, SMO, Klasifikasi

Image Classification Using Random Forest Method and Sequential Minimal Optimazation (SMO)

Abstract

Classification aims to separate the types of categories and labels that have been made. This research tries to classify the flower image using the segmentation process. We use this segmentation to remove background noise. Apart from that to get the image features that we classify. The features we use are eccentricity, perimeter, metric, and area. After that, we process it to get a grade. We used 120 images from a dataset of 17 flowers with 2 types of flowers in white and yellow. Then we divided it into training data and test data using the Random forest and SMO classification methods processed using 10-fold cross-validation and 66% split scenario. The result is that the random forest method has the best accuracy with 2 different scenarios compared to the SMO method.

Keywords: Random Forest, SMO, Classification.

I. PENDAHULUAN

Klasifikasi merupakan Teknik yang paling umum diterapkan pada data mining. Proses klasifikasi data melibatkan learning dan klasifikasi. Dalam belajar (learning) data pelatihan (training) dianalisis dengan algoritma klasifikasi. Dalam proses klasifikasi pengujian data dilakukan dengan menggunakan perkiraan akurasi dari aturan klasifikasi[1].

Sedangkan klasifikasi gambar merupakan tugas untuk menetapkan label ke gambar yang telah kita tentukan kategorinya untuk dijadikan informasi [2]. Secara praktis, berarti tugas klasifikasi citra adalah untuk menganalisis input citra dan mengembalikan label untuk dikategorikan. Label pada suatu citra sudah ditentukan sebelumnya. Sebagai contoh, kita misalkan bahwa dataset yang memungkinkan menggunakan kategori = {kucing, anjing, kuda, panda, burung}. Kemudian didalam dataset tersebut

akan muncul gambar-gambar yang meliputi kategori tersebut untuk dimasukan kedalam sistem klasifikasi.

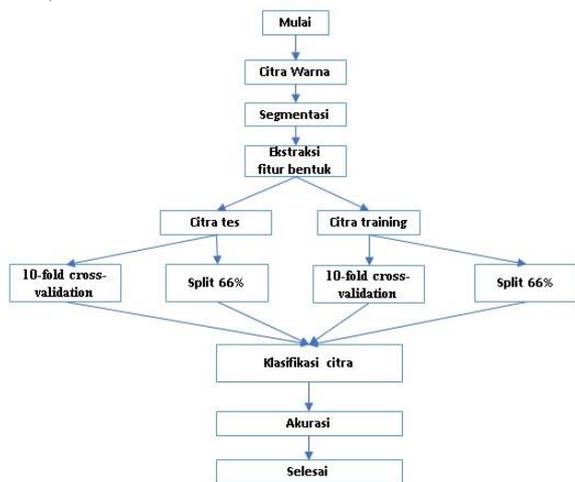
Klasifikasi bertujuan untuk memisahkan antara jenis-jenis kategori dan label yang telah dibuat. Didalam penelitian ini kami mencoba mengklasifikasi citra bunga dengan menggunakan proses segmentasi. Segmentasi kami lakukan menggunakan metode otsu [3] untuk meningkatkan tingkat akurasi. Segmentasi ini kami gunakan untuk menghilangkan noise yang ada didalam background. Kemudian kami bagi menjadi data training dan data test dengan pembagian 10-fold cross-validation dan 66% split.

Penelitian sebelumnya melakukan klasifikasi bunga menggunakan klasifikasi naïve bayes [4] dan mendapatkan nilai terbaik sebesar 97%. Lalu Saini[5] melakukan identifikasi pada daun menggunakan klasifikasi Random Forest dan SMO mendapatkan nilai terbaik sebesar 80% dengan total sample 65 gambar.

Pada penelitian ini kami mencoba melakukan klasifikasi Bunga dari dataset 17Flower dengan jumlah sample sebanyak 120 citra dengan menggunakan deteksi canny edge [6] karena dinilai memiliki deteksi paling baik dari deteksi tepi lainnya. Sehingga dapat mempengaruhi tingkat akurasi didalam proses klasifikasi citra.

II. METODOLOGI

Berikut adalah metode penelitian usulan kami sesuai dengan gambar1. Kami menggunakan citra warna dengan jumlah sample 120 gambar dan menggunakan klasifikasi random forest dan Sequential Minimal Optimization (SMO)



Gambar 1. Metode Penelitian

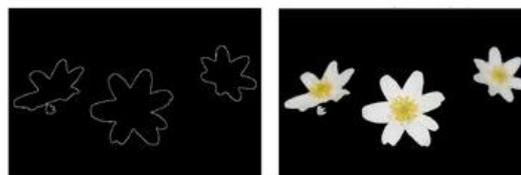
A. Segmentasi

Segmentasi citra adalah proses pengolahan citra yang bertujuan memisahkan wilayah (region) objek dengan wilayah latar belakang agar objek mudah dianalisis dalam rangka mengenali objek yang banyak melibatkan persepsi visual.[7]. Tujuan untuk segmentasi warna adalah mencari informasi warna untuk mendefinisikan wilayah[8].

B. Canny Edge

Canny edge adalah lokasi tepi yang cocok administrator menggunakan perhitungan multi tahap untuk mengenali berbagai sisi dalam gambar. Poin utama dari perhitungan ini adalah untuk menemukan perhitungan ideal untuk deteksi tepi. Menurut selvakumar dalam keadaan ini, deteksi tepi dikatakan "ideal" berarti:

- 1) Penemuan Hebat: Perhitungan harus memiliki jumlah tepi aktual yang sama pada citra seperti yang diharapkan dalam kondisi citra (background) tersebut.
- 2) Pengekangan besar: tepi yang dibatasi harus sedekat mungkin dengan tepi pada gambar aktual.
- 3) Jawaban tidak penting: batas tertentu pada citra hanya boleh diperiksa sekali dan, jika mungkin citra memiliki banyak noise seharusnya tidak membuat tepi yang salah [6]. Maka sebelum citra ini dieksekusi kedalam canny edge dilakukan pembersihan noise terlebih dahulu.



a. canny edge

b. hasil cat and mask

Gambar 2. Hasil canny edge

C. Ekstraksi Citra

Beberapa penelitian menerapkan adanya ekstraksi fitur dengan beberapa model untuk meningkatkan tingkat akurasi yang baik pada saat proses klasifikasi [9]. Seperti penelitian yang dilakukan oleh Amira [10], melakukan ekstraksi fitur untuk mendapatkan nilai yang baik didalam proses klasifikasi menggunakan SVM. Didalam pengambilan fiturnya Amira melakukan proses training dan uji, setelah itu menggunakan segmentasi untuk mendapatkan fitur warna, texture, dan bentuk.

D. Klasifikasi

Klasifikasi adalah cara untuk mengidentifikasi Teknik pengelompokan dari banyak data yang diberikan dan bergantung pada target atribut atau output, seluruh kumpulan data bisa memenuhi syarat menjadi anggota class. Teknik ini membantu untuk mengidentifikasi perilaku pola data. Singkatnya ini merupakan mekanisme diskriminasi data[11].

Klasifikasi random forest dilakukan melalui pembentukan pohon (tree) dengan melakukan training pada sampel data yang dimiliki. Random Forest dalam melakukan klasifikasi, menggunakan cara hasil voting keputusan terbanyak berdasarkan dari pohon yang telah dibentuk[12].

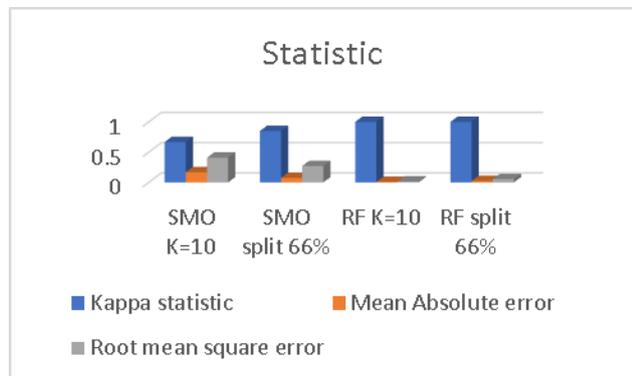
Metode klasifikasi yang digunakan adalah metode random forest dan *support vector machine*. Sebelum dilakukan klasifikasi, dataset dibagi ke dalam 2 kategori yaitu data untuk *training* dan data untuk *testing*. Data dipisahkan menggunakan metode *K-fold cross validation* dengan $k = 10$ dan split 66% menjadi data *training* dan data *testing*. Pemisahan data ini dilakukan untuk mengetahui performa dari sistem[13].

III. HASIL DAN PEMBAHASAN

Didalam penelitian ini kami mempunyai 120 gambar yang sudah diolah dengan ekstraksi fitur. Pengambilan fitur adanya fitur bentuk yang terdiri dari eccentricity, metric, perimeter dan area. Pengambilan fitur ini dengan menggunakan tool matlab sebagai alat bantu untuk pengolahan citra. Sehingga menghasilkan data dengan ukuran 120×4 . Kemudian kami bagi data tersebut menjadi data training dan data testing. . Data dipisahkan menggunakan metode *K-fold cross validation* dengan $k = 10$ dan split 66% menjadi data *training* dan data *testing*. Dengan menggunakan klasifikasi metode random forest dan SMO. Dalam proses klasifikasi ini kami menggunakan *weka tools*.

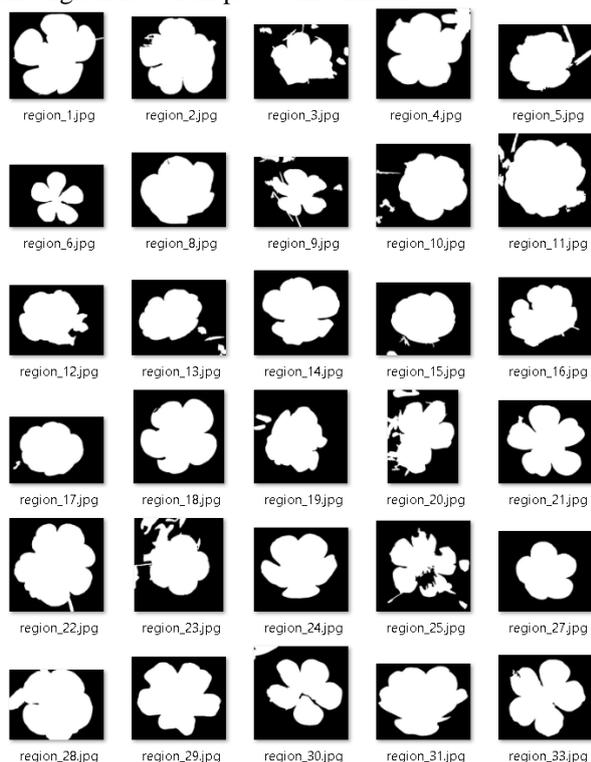
Gambar 3. Menunjukkan perbedaan statistic antara klasifikasi Random Forest (RF) dengan SMO. Dapat

dilihat untuk kappa statistic pada random forest bernilai 1 sedangkan SMO di bawah nilai 1. Kemudian untuk nilai errornya RF memiliki nilai lebih rendah sebesar 0.0065 untuk model klasifikasi K=10 dan 0.0224 dengan model klasifikasi split 66% dibandingkan dengan metode SMO.



Gambar 3. statistic

Gambar 4 merupakan hasil dari proses segmentasi yang akan di ambil fitur bentuk yaitu eccentricity, perimeter, metric dan area. Dari gambar ini akan dihasilkan nilai-nilai yang di sebut hasil ekstrasi sehingga dapat di perhitungkan ke dalam proses klasifikasi.



Gambar 4. Hasil pembersihan noise

Maka dari perhitungan menggunakan weka tools akurasi yang di dapat untuk setiap metode dapat kita lihat pada table 1. Random forest mendapatkan nilai sempurna untuk klasifikasi ini sebesar 100%.

TABEL I
AKURASI

	K=10	SPLIT 66%
SMO	83.33%	92.68%
RANDOM FOREST	100%	100%

IV. KESIMPULAN

Berdasarkan hasil penelitian kami, maka metode klasifikasi menggunakan Random Forest mendapatkan hasil yang lebih baik yaitu sebesar 100% untuk kedua scenario k=10 dan split 66%, hal ini disebabkan karena adanya proses segmentasi untuk membersihkan noise yang ada sehingga area terpenting saja yang terhitung oleh sistem. Sedangkan SMO tidak berbeda jauh dengan metode Random Forest untuk scenario split 66% mendapatkan nilai akurasi sebesar 92.68%.

UCAPAN TERIMA KASIH

Peneliti mengucapkan terima kasih kepada LPPM Universitas Pamulang serta Program studi Informatika Universitas Pamulang yang telah mendukung kami untuk melakukan penelitian ini.

DAFTAR PUSTAKA

- [1] Agus Nur Khormarudin, "Teknik Data Mining : Algoritma K-Means Clustering," *IlmuKomputer.Com*, pp. 1–12, 2016.
- [2] A. Rosebrock, "Deep Learning for Computer Vision with Python," *PyImageSearch*, vol. 53, p. 160, 2017.
- [3] P. Rosyani, "Deteksi Citra Bunga Menggunakan Analisis Segmentasi Fuzzy C-Means dan Otsu Threshold," vol. 20, no. 1, 2020.
- [4] P. Rosyani and O. Hariansyah, "Pengenalan Citra Bunga Menggunakan Segmentasi Otsu Threshold dan Naïve Bayes," pp. 1–7, 2020.
- [5] M. S. Sainin and R. Alfred, "Feature selection for Malaysian medicinal plant leaf shape identification and classification," *2014 Int. Conf. Comput. Sci. Technol. ICCST 2014*, vol. 2014, 2014.
- [6] P. Selvakumar and S. Hari Ganesh, "Tamil Character Recognition Using Canny Edge Detection Algorithm," *Proc. - 2nd World Congr. Comput. Commun. Technol. WCCCT 2017*, pp. 250–254, 2017.
- [7] D. Anggraeni, V. A. Putri, S. F. Al-Uswah, M. F. H. Tulloh, and M. S. Zuhri, "Segementasi citra digital ikan menggunakan metode thresholding dan k-means," pp. 1–10, 2009.
- [8] P. Rosyani, M. Taufik, A. A. Waskita, and D. H. Apriyanti, "Comparison of color model for flower recognition," in *2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)*, 2019, pp. 10–14.
- [9] J. C. Kavitha and A. Suruliandi, "Texture and color feature extraction for classification of melanoma using SVM," *2016 Int. Conf. Comput. Technol. Intell. Data Eng. ICCTIDE 2016*, 2016.
- [10] A. B. Mabrouk, A. Najjar, and E. Zagrouba, "Image flower recognition based on a new method for color feature extraction," *VISAPP 2014 - Proc. 9th Int. Conf. Comput. Vis. Theory Appl.*, vol. 2, 2014.
- [11] B. Derviş, *濟無No Title No Title*, vol. 53, no. 9, 2013.
- [12] M. Wahyu, B. Santoso, R. C. Wihandika, and M. A. Rahman, "Ekstraksi Ciri untuk Klasifikasi Jenis Kelamin berbasis Citra Wajah menggunakan Metode Compass Local Binary Patterns," vol. 3, no. 11, 2019.
- [13] A. Mustofa, H. Tjandrasa, and B. Amaliah, "Deteksi Penyakit Glaukoma pada Citra Fundus Retina Mata Menggunakan Adaptive Thresholding dan Support Vector Machine," *J. Tek. ITS*, vol. 5, no. 2, 2016.