

PENGARUH PENGGUNAAN *INFORMATION GAIN* UNTUK SELEKSI FITUR CITRA TANAH DALAM RANGKA MENILAI KESESUAIAN LAHAN PADA TANAMAN CENGKEH

Danang Aji Bimantoro^[1], Shofwatul 'Uyun^[2]

Teknik Informatika, Universitas Islam Negeri Sunan Kalijaga Yogyakarta

Jl. Marsda Adisucipto Yogyakarta 55281

E-mail :danang.aji.bimantoro@gmail.com^[1] , shofwatul.uyun@uin-suka.ac.id^[2]

ABSTRACT

Clove is a commodity that has a high economic value, so it needs some additional field to increase production. Proper field determined by many factors, one of which is the characteristics of the soil. Soil suitability for clove plants can be distinguished visually using images based on color and texture features. The parameters used from the extraction results of both features are the average value of (red, green, blue) and (mean, variance, kurtosis, skewness, entropy). The use of all these parameters is not positively correlated with the accuracy of soil suitability assessment for clove plants. Therefore, it is necessary to make feature selection process to get parameters that have influence in determination of soil recognition result. This study analyzes the influence of the use of information gain for the selection of soil image features in order to assess the suitability of soil in clove plants. The amount of data used in this study is 50 images that are divided into two appropriate land images and not suitable for clove plants. This research stage begins with image acquisition, pre-processing of image, feature extraction, feature selection, clustering followed by analysis. The results of the analysis show that the use of feature results from the feature selection process using information gain proved to increase the accuracy value. The test results show the accuracy level of feature use without selection process is only 50%, while the selected feature of the selection using information gain with threshold value of 0.7 increased to 88%.

Keyword: image processing, information gain, fuzzy c-means, soil, clove

ABSTRAK

Cengkeh merupakan komoditas perkebunan yang memiliki nilai ekonomis cukup tinggi sehingga perlu penambahan lahan untuk meningkatkan produksinya. Penentuan lahan baru yang tepat perlu mempertimbangkan banyak faktor, salah satunya adalah karakteristik sifat tanah itu sendiri. Kesesuaian lahan untuk tanaman cengkeh dapat dibedakan secara visual menggunakan citra berdasarkan fitur warna dan tekstur. Parameter yang digunakan dari hasil ekstraksi kedua fitur tersebut adalah nilai rata-rata dari (*red, green, blue*) serta (*mean, variance, kurtosis, skewness, entropy*). Penggunaan seluruh parameter tersebut tidak berkorelasi positif dengan hasil akurasi penilaian kesesuaian lahan untuk tanaman cengkeh. Oleh karena itu perlu dilakukan proses seleksi fitur untuk mendapatkan parameter yang memiliki pengaruh dalam penentuan hasil pengenalan lahan. Pada penelitian ini dilakukan analisis pengaruh penggunaan *information gain* untuk seleksi fitur citra tanah dalam rangka menilai kesesuaian lahan pada tanaman cengkeh. Jumlah data yang digunakan pada penelitian ini adalah 50 citra yang terbagi menjadi dua citra tanah yang sesuai dan tidak sesuai untuk tanaman cengkeh. Tahapan penelitian ini diawali dengan akuisisi citra, pra pengolahan terhadap citra, ekstraksi fitur, seleksi fitur, clustering dilanjutkan dengan analisis. Hasil analisis menunjukkan penggunaan fitur hasil dari proses seleksi fitur menggunakan *information gain* terbukti mampu meningkatkan nilai akurasi. Hasil pengujian menunjukkan tingkat akurasi penggunaan fitur tanpa proses seleksi hanya 50%, sedangkan fitur terpilih dari hasil seleksi menggunakan *information gain* dengan nilai *threshold* 0,7 naik menjadi 88%.

Kata Kunci: pengolahan citra digital, *information gain*, fuzzy c-means, tanah, cengkeh

1. PENDAHULUAN

Tanaman cengkeh merupakan salah satu tanaman asli Indonesia yang memiliki nilai ekonomis cukup tinggi. Tanaman ini berasal dari kepulauan Maluku, seperti: Tidore, Ternate, Mutir, dan sebagainya. Sebagian besar cengkeh saat ini dimanfaatkan sebagai bahan dasar dari rokok kretek, namun selain itu cengkeh juga dimanfaatkan sebagai bahan obat-obatan tradisional, obat-obatan modern, bahan baku pembuatan *vanillin*, bumbu dapur dan bahan wangi-wangian. Selain cengkehnya, daun dan juga tangkai/gagangnya juga dapat didistilasi untuk diambil minyaknya walaupun mutunya tidak sebaik cengkeh. Selain itu batang yang sudah kering/mati dari tanaman cengkeh dapat dimanfaatkan sebagai perkakas rumah tangga dan juga kayu bakar. Harga bunga cengkeh kering per kilogramnya saat ini kurang lebih Rp 90.000,- di wilayah Desa Gerbosari, Kecamatan Samigaluh, Kabupaten Kulon Progo, D.I. Yogyakarta. (AAK, 1981)

Nilai ekonomis tanaman cengkeh yang cukup tinggi membuat masyarakat Desa Gerbosari, Kecamatan Samigaluh, Kabupaten Kulon Progo berusaha untuk melakukan penanaman tanaman cengkeh dilahan yang lain (lahan yang belum ditanami cengkeh). Namun dari beberapa percobaan penanaman terdapat sebagian tanaman cengkeh yang tidak dapat bertahan hidup. Penyebab tanaman cengkeh yang ditanam tersebut tidak mampu bertahan hidup dipengaruhi beberapa faktor, yaitu ketinggian area penanaman, curah hujan, suhu udara, kelembaban udara, tanah dan perawatan.

Tanah merupakan salah satu parameter yang mempengaruhi tingkat kehidupan dari tanaman cengkeh. Tanah yang sesuai untuk tanaman cengkeh memiliki ciri-ciri memiliki nilai pH 5,5 – 6,5, berwarna coklat gelap dan memiliki tekstur gembur sehingga mampu mengikat air. Dari hasil wawancara dengan petani cengkeh, tingkat kematian tanaman cengkeh jika ditanam di tanah yang tidak sesuai mencapai 100%, karena karakteristik tanah tersebut tidak mampu memenuhi kebutuhan dari tanaman cengkeh, terutama air. (AAK, 1981)

Pada penelitian ini, hanya digunakan parameter citra tanah untuk menilai kesesuaian lahan pada tanaman cengkeh karena untuk parameter lain seperti ketinggian area lahan, curah hujan, suhu udara dan kelembaban untuk di wilayah Desa Gerbosari, Kecamatan Samigaluh, Kabupaten Kulon Progo sudah memenuhi kriteria penanaman tanaman cengkeh. Sedangkan digunakannya citra tanah dalam penelitian ini didasarkan pada keilmuan pakar yang dapat membedakan tanah yang sesuai dan tidak sesuai untuk tanaman cengkeh secara visual, terutama dari warna dan tekstur tanah.

Citra tanah yang digunakan nantinya akan melalui proses ekstraksi fitur, yaitu proses untuk mendapatkan data karakteristik dari citra tersebut. Pada penelitian ini digunakan fitur warna dan tekstur sesuai dengan keilmuan pakar yang dapat membedakan kesesuaian tanah dari warna dan tekstur tanah. Fitur warna yang digunakan dalam penelitian ini adalah rata-rata nilai *red*, nilai *green* dan nilai *blue* dari keseluruhan piksel citra. Parameter *red*, *green* dan *blue* digunakan karena pada dasarnya sebuah citra digital pasti akan memiliki parameter tersebut tanpa perlu adanya konversi ke model warna lainnya. Sedangkan fitur tekstur yang digunakan adalah ekstraksi fitur orde pertama yang menghasilkan parameter *mean*, *variance*, *kurtosis*, *skewness* dan *entropy*. Ekstraksi fitur orde pertama digunakan karena dalam proses ekstraksi tidak memakan biaya komputasi yang terlalu besar, hal ini disebabkan tekstur orde pertama dihitung berdasarkan nilai histogram citra tanah setelah diubah ke dalam bentuk citra *grayscale*.

Pada penelitian ini, pengenalan pola akan digunakan dalam proses penilaian kesesuaian lahan pada tanaman cengkeh. Pengenalan pola merupakan salah satu kemampuan dasar yang sudah secara alami dimiliki manusia dan merupakan hal yang sangat mendasar dari kecerdasan manusia (Yuan & Klir, 1995). Namun dalam penelitian ini, pengenalan pola diartikan sebagai salah satu cabang kecerdasan buatan yang menitikberatkan pada metode pengklasifikasian objek kedalam kelas-kelas tertentu. Algoritma pengenalan pola yang akan digunakan adalah *Fuzzy C-Means*. Algoritma ini termasuk kedalam algoritma *clustering* yang dalam proses kerjanya akan mengelompokkan citra tanah kedalam kelompok citra tanah yang sesuai dan tidak sesuai untuk tanaman cengkeh berdasarkan fitur-fitur yang dimiliki oleh citra tanah. Namun fitur hasil ekstraksi belum tentu dapat memberikan hasil *clustering* terbaik. Seleksi fitur perlu dilakukan agar fitur yang digunakan dalam proses *clustering* merupakan fitur terbaik. *Information gain* merupakan salah satu metode untuk seleksi fitur, yang akan memberikan nilai

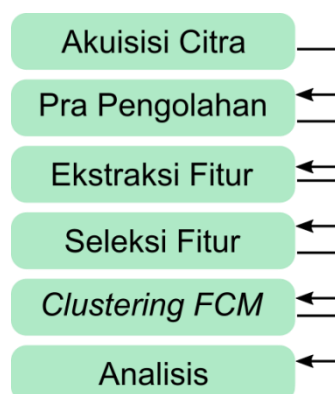
untuk setiap fitur dan dapat dilakukan eliminasi fitur yang memiliki nilai rendah. Seleksi fitur dilakukan karena terdapat kemungkinan parameter fitur yang dihasilkan dari proses ekstraksi fitur memberikan akurasi *clustering* rendah.

Penelitian yang dilakukan oleh Sari (2016) tentang implementasi seleksi fitur *information gain* pada algoritma machine learning untuk prediksi performa akademik menghasilkan kesimpulan bahwa seleksi fitur *information gain* memberikan pengaruh terhadap hasil prediksi. Penelitian yang dilakukan oleh Wahyuni (2016) tentang penerapan metode seleksi fitur untuk meningkatkan hasil diagnosis kanker payudara menunjukkan bahwa seleksi fitur mampu memberikan akurasi diagnosis yang lebih baik dan ukuran dataset yang lebih kecil, penelitian ini menggunakan seleksi fitur *rough set* dan *F-score* dengan algoritma untuk diagnosis *SMO* (*Sequential Minimal Optimization*), *Naive Bayes*, *Multi Layer Perceptron* dan *C.45*. Pada penelitian yang dilakukan oleh Maulana dan Chasanah (2016) tentang peningkatan performa algoritma *C4.5* dengan seleksi fitur *information gain* untuk klasifikasi persetujuan kredit menunjukkan bahwa seleksi fitur *information gain* mampu meningkatkan akurasi klasifikasi dari 94,12% menjadi 95,69%.

Pada penelitian ini akan dilakukan analisa pengaruh penggunaan *information gain* untuk seleksi fitur citra tanah dalam rangka menilai kesesuaian lahan pada tanaman cengkeh.

2. METODE PENELITIAN

Metodologi penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Metodologi Penelitian

2.1. Akuisisi Citra

Akuisisi citra merupakan tahapan pengumpulan data citra tanah yang akan digunakan dalam penelitian ini. Citra tanah yang dikumpulkan merupakan citra tanah yang sesuai dan tidak sesuai untuk tanaman cengkeh. Proses akuisisi citra dilakukan di wilayah Desa Gerbosari, Kecamatan Samigaluh, Kabupaten Kulon Progo, D.I. Yogyakarta. Proses akuisisi citra dilakukan menggunakan kamera DSLR Nikon D5000 dengan standarisasi pengaturan kamera dan jarak pengambilan citra agar citra tanah yang didapatkan kualitas yang sama. Penentuan titik pengambilan citra tanah didasarkan pada pakar.

2.2. Pra Pengolahan

Pra pengolahan pada hasil akuisisi citra adalah proses *cropping*. Proses ini dilakukan untuk mendapatkan ukuran citra yang tidak terlalu besar, sehingga memudahkan dalam tahapan selanjutnya. Standar *cropping* yang digunakan adalah 1:1 dengan menggunakan perangkat lunak microsoft office picture manager.

2.3. Ekstraksi Fitur

Tahapan ekstraksi fitur dilakukan untuk mendapatkan data ciri atau fitur yang dimiliki oleh citra tanah. Ekstraksi fitur yang digunakan dalam penelitian ini adalah ekstraksi ciri warna (*mean red, mean green, mean blue*) dan ekstraksi ciri tekstur orde pertama (*mean, variance, kurtosis, skewness, entropy*). Penggunaan rata-rata nilai red, nilai green dan nilai blue dari keseluruhan piksel citra tanah pada penelitian ini dikarenakan citra digital yang didapatkan dari perangkat kamera menggunakan model warna RGB, sehingga tidak perlu dilakukan konversi ke model warna yang lainnya. Sedangkan pemilihan tekstur orde pertama dikarenakan dalam proses ekstraksi tidak akan memakan biaya komputasi yang terlalu besar, hal ini disebabkan tekstur orde pertama dihitung berdasarkan nilai histogram citra tanah.

2.4. Seleksi Fitur

Algoritma seleksi fitur yang digunakan pada penelitian ini adalah *information gain*. *Information gain* merupakan salah satu algoritma untuk seleksi fitur yang banyak dipakai untuk menentukan batas dari kepentingan sebuah atribut. *Information gain* merupakan metode yang digunakan dalam algoritma *machine learning decision tree*. *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. Peran *Information gain* pada *decision tree* adalah untuk menyusun struktur *tree*. Hanya saja dalam penyusunan jaringan *tree* tidak dilakukan perhitungan jarak antara entropy keseluruhan data. Nilai *information gain* diperoleh dari nilai *entropy* sebelum pemisahan dikurangi dengan nilai *entropy* setelah pemisahan. Pengukuran nilai ini hanya digunakan sebagai tahapan awal untuk penentuan atribut yang nantinya akan digunakan atau dibuang. Atribut yang memenuhi kriteria pembobotan yang nantinya akan digunakan dalam proses pengenalan pola.

Pemilihan fitur dengan *information gain* dilakukan dalam 3 tahapan, yaitu (Maulana & Karomi, 2015)

1. Menghitung nilai *information gain* untuk setiap atribut dalam dataset asli.
2. Menentukan batas (*threshold*) yang diinginkan. Hal ini akan memungkinkan atribut yang berbobot sama dengan batas atau lebih besar akan dipertahankan atribut yang berada dibawah batas.
3. Dataset diperbaiki dengan pengurangan atribut.

Perhitungan nilai *information gain* adalah sebagai berikut :

1. Menghitung *entropy* sebelum pemisahan

$$Info(D) = \sum_{i=1}^c p_i \log_2(p_i) \quad (1)$$

Dimana :

c = Jumlah kelas data

p_i = jumlah sampel data untuk kelas i

2. Menghitung *entropy* setelah pemisahan atribut

$$Info_A = \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j) \quad (2)$$

Dimana :

A = Atribut

$|D|$ = Jumlah seluruh sampel data

$|D_j|$ = jumlah sampel data untuk nilai j

v = nilai yang mungkin untuk atribut A

3. Menghitung nilai *information gain*

$$Gain(A) = |Info(D) - Info_A(D)| \quad (3)$$

2.5. Clustering

Algoritma *clustering* yang digunakan dalam penelitian ini adalah *Fuzzy C-Means*. Data yang diperoleh dari tahapan seleksi fitur akan digunakan untuk mengelompokkan data tersebut kedalam kelompok-kelompok tertentu.

Tahapan awal algoritma *Fuzzy C-Means* adalah menentukan parameter yang dibutuhkan dalam proses *clustering*, yaitu: jumlah *cluster*, pemangkat, maksimum iterasi, *error* terkecil yang diharapkan, fungsi objektif awal dan iterasi awal. Kemudian setelah parameter ditentukan, tahapan berikutnya adalah membangkitkan matrik partisi awal. Tahapan selanjutnya adalah mencari pusat *cluster*. Kemudian dari pusat *cluster* yang sudah didapatkan dihitung fungsi objektif dan dilanjutkan dengan memperbaharui nilai dari matrik partisi. Tahapan berikutnya melakukan pengecekan terhadap kondisi perulangan. Jika iterasi kurang dari maksimum iterasi atau fungsi objektif dikurangi fungsi objektif iterasi sebelumnya lebih dari *error* terkecil, maka proses perulangan akan kembali ke tahapan pencarian pusat *cluster*, sebaliknya jika kondisi tersebut tidak terpenuhi, maka proses perulangan akan dihentikan dan pusat *cluster* pada iterasi terakhir merupakan pusat *cluster* yang digunakan sekaligus matrik partisi dari iterasi terakhir merupakan hasil pengelompokan data.

2.6. Analisis

Proses analisis dilakukan untuk mendapatkan informasi dari penerapan seleksi fitur *information gain* pada proses *clustering Fuzzy C-Means*

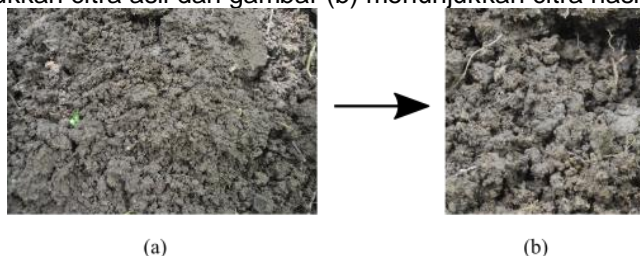
3. HASIL DAN PEMBAHASAN

Tahapan awal penelitian ini adalah akuisisi citra, tahapan ini menghasilkan citra tanah sejumlah 50 citra dengan pembagian 25 citra tanah yang sesuai untuk tanaman cengkeh dan 25 citra tanah yang tidak sesuai berdasarkan pakar. Citra yang dihasilkan dari tahapan ini berukuran 4288 x 2848 piksel dengan format *JPG. Gambar 2 merupakan beberapa hasil proses akuisisi citra tanah. Gambar (a) adalah citra tanah yang sesuai untuk tanaman cengkeh berdasarkan pakar, sedangkan gambar (b) adalah citra yang tidak sesuai untuk tanaman cengkeh.



Gambar 2. Hasil Akuisisi Citra

Hasil akuisisi citra memiliki ukuran yang terlalu besar, sehingga perlu dilakukan pra pengolahan terhadap citra. *Cropping* dilakukan dengan skala 1:1. Pada penelitian ini ukuran citra dipotong menjadi 1000 x 1000 piksel, ukuran tersebut tidak terlalu besar dan tidak terlalu kecil sehingga masih dapat dilakukan ekstraksi tekstur. Gambar 3 menunjukkan tahapan pra pengolahan. Gambar (a) menunjukkan citra asli dan gambar (b) menunjukkan citra hasil *cropping*.



Gambar 3. Citra Hasil *Cropping*

Proses selanjutnya adalah melakukan ekstraksi fitur, fitur yang digunakan adalah warna (mean red, mean green, mean blue) dan tekstur orde pertama (*mean, variance, kurtosis, skewness, entropy*). Hasil ekstraksi fitur untuk beberapa citra dapat dilihat pada Tabel 1.

Tabel 1 . Hasil Ekstraksi Fitur

No	Citra	R	G	B	Mean	Var	Skew	Kurt	Ent
1	TS_1.JPG	106	104	105	105	1234	0	-1	-7
2	TS_2.JPG	72	70	70	71	814	0	0	-7
3	TTS_1.JPG	176	142	108	142	406	1	2	-6
4	TTS_2.JPG	172	132	86	130	199	0	12	-5

Data hasil ekstraksi fitur akan dilakukan seleksi dengan *information gain*. Tahapan ini akan menghasilkan nilai untuk setiap masing masing fitur yang dapat digunakan sebagai acuan dalam proses eliminasi berdasarkan batasan (*threshold*) yang ditentukan. Berikut ini adalah pseudocode perhitungan nilai *information gain* dengan menerapkan persamaan (1), (2) dan (3).

```
//menghitung entropy keseluruhan (entropy awal)
Entropy = - (T[1] /total)*log2(T[1] /total) - (T[2] /total)*log2(T[2] /total);

For j=0, ... jml_fitur do
  Ent_Ftr(Fiturj) = Di/total* entropy(Di) +...+ Di+n/total*
  entropy(Di+n) ;
  //n merupakan banyak data yang mungkin ada pada fitur

  Info(fiturj) = | Entropy - Ent_Ftr(fiturj) |;
end
```

Hasil perhitungan nilai *information gain* dapat dilihat pada Tabel 2.

Tabel 2. Hasil Perhitungan Information Gain

No.	Fitur	Information Gain	Jumlah nilai yang mungkin muncul
1	Mean red	1	48 nilai
2	Mean green	1	40 nilai
3	Mean blue	0,655098	36 nilai
4	Mean	1	38 nilai
5	Variance	1	50 nilai
6	Skewness	0,313076	3 nilai

Tabel 2. Lanjutan Perhitungan Information Gain

No.	Fitur	Information Gain	Jumlah nilai yang mungkin muncul
7	Kurtosis	0,73672	5 nilai
8	Entropy	0,156513	3 nilai

Berdasarkan hasil perhitungan nilai *information gain* pada Tabel 2, dapat dilakukan beberapa kemungkinan nilai *threshold* untuk melakukan eliminasi fitur berdasarkan nilai *information gain*, yaitu 0,2, 0,5, 0,7 dan 1. Selain itu pada penelitian ini akan dilakukan percobaan *clustering* dengan seluruh fitur atau nilai *threshold* adalah 0.

Setelah melakukan ekstraksi fitur, tahapan selanjutnya adalah melakukan *clustering* dengan algoritma fuzzy c-means berdasarkan fitur yang didapatkan dari hasil seleksi fitur *information gain*. Proses *clustering* dilakukan dengan parameter yang sama pada setiap percobaan, yaitu

menggunakan nilai pemangkat (w) 2, dan maksimum iterasi sampai 1000 iterasi. Percobaan pertama dilakukan dengan seluruh fitur atau nilai *threshold* adalah 0. Hasil clustering dapat dilihat pada Tabel 3.

Tabel 3. Hasil Clustering Semua Fitur

No.	Fitur	Pusat Cluster		Iterasi	Akurasi
		Cluster 1	Cluster 2		
1	<i>Mean red</i>	137.92	3.1057467123041e19	1000	50%
2	<i>Mean green</i>	117.22	2.6396144838768e19		
3	<i>Mean blue</i>	91.44	2.0590884525311e19		
4	<i>Mean</i>	115.54	2.6017834624392e19		
5	<i>Variance</i>	1183.66	2.6654206449288e20		
6	<i>Skewness</i>	-0.34	-7.6562781480814e16		
7	<i>Kurtosis</i>	0.3	6.7555395424248e16		
8	<i>Entropy</i>	-6.84	-1.5402630156729e18		

Percobaan kedua dilakukan dengan menggunakan fitur dengan nilai information lebih dari atau sama dengan 0,2 atau dengan menggunakan nilai *threshold*=0,2. Hasil penggunaan *threshold* 0,2 dapat dilihat pada Tabel 4.

Tabel 4. Hasil Clustering dengan Threshold 0,2

No.	Fitur	Pusat Cluster		Iterasi	Akurasi
		Cluster 1	Cluster 2		
1	<i>Mean red</i>	112.4035712941	373.7280710249	21	80%
2	<i>Mean green</i>	101.7206789732	318.69053153352		
3	<i>Mean blue</i>	88.48530262372	250.52201111679		
4	<i>Mean</i>	100.8810378886	314.34838235313		
5	<i>Variance</i>	1371.665826306	3257.4703934186		
6	<i>Skewness</i>	0.035458026909	-0.91674936120723		
7	<i>Kurtosis</i>	-0.21061670338	0.78078759896003		

Percobaan berikutnya dilakukan dengan nilai *threshold* 0,5. Hasil clustering pada percobaan ini dapat dilihat pada Tabel 5.

Tabel 5. Hasil Clustering dengan Threshold 0,5

No.	Fitur	Pusat Cluster		Iterasi	Akurasi
		Cluster 1	Cluster 2		
1	<i>Mean red</i>	125.7145992188	506.11582871075	39	78%
2	<i>Mean green</i>	109.7267043890	428.6627985707		
3	<i>Mean blue</i>	88.45692050576	330.36149803562		
4	<i>Mean</i>	107.9820308142	421.78624582164		
5	<i>Variance</i>	1360.014291521	4269.0591967415		
6	<i>Kurtosis</i>	-0.06952936612	1.1544353326044		

Percobaan keempat dilakukan dengan menggunakan nilai *threshold* 0,7. Hasil percobaan ini dapat dilihat pada Tabel 6.

Tabel 6. Hasil Clustering dengan Threshold 0,7

No.	Fitur	Pusat Cluster		Iterasi	Akurasi
		Cluster 1	Cluster 2		
1	<i>Mean red</i>	105.5128956658	250.73468086545	32	88%
2	<i>Mean green</i>	97.898558133573	214.58859975283		
3	<i>Mean</i>	97.503683169494	211.3635743349		
4	<i>Variance</i>	1464.7389045799	2305.5077759535		
5	<i>Kurtosis</i>	-0.603114975187	0.14617788012236		

Percobaan Terakhir dilakukan dengan menggunakan nilai *threshold* 1. Hasil penggunaan nilai *threshold* 1 dapat dilihat pada Tabel 7.

Tabel 7. Hasil Clustering dengan Threshold 1

No.	Fitur	Pusat Cluster		Iterasi	Akurasi
		Cluster 1	Cluster 2		
1	<i>Mean red</i>	137.92	4.5698699230164e73	1000	50%
2	<i>Mean green</i>	117.22	3.8839918240718e73		
3	<i>Mean</i>	115.54	3.8283263551719e73		
4	<i>Variance</i>	1183.66	3.9219636260713e74		

Berdasarkan percobaan yang sudah dilakukan, dapat diketahui bahwa seleksi fitur *information gain* sangat mempengaruhi proses dan hasil *clustering*. Penggunaan nilai *threshold* 0,7 memberikan hasil *clustering* terbaik, yaitu menghasilkan akurasi sebesar 88%. Namun dari percobaan yang sudah dilakukan juga diketahui bahwa nilai *information gain* terbaik belum tentu memberikan hasil *clustering* terbaik, terbukti pada percobaan terakhir dengan penggunaan nilai *threshold* 1, akurasi yang dihasilkan hanya 50%. Pada Tabel 2 dapat diketahui bahwa nilai *information gain* sangat dipengaruhi oleh jumlah nilai yang mungkin muncul pada setiap fitur, semakin banyak kemungkinan nilai yang muncul maka semakin tinggi kemungkinan nilai *information gain* dari fitur tersebut. Suatu fitur akan memiliki nilai *information gain* yang baik apabila nilai-nilai yang ada pada fitur tersebut memiliki kecenderungan karakteristik yang berbeda untuk setiap kelompok data. Pada Tabel 8 dapat dilihat analisis pengaruh seleksi fitur *information gain* terhadap *clustering* menggunakan algoritma *fuzzy c-means*.

Tabel 8. Analisa Pengaruh Seleksi Fitur Terhadap Hasil Clustering

Threshold	Fitur	Clustering	
		Akurasi	Iterasi
Tanpa seleksi fitur	semua fitur	50%	1000
0,2	<i>mean red, mean green, mean blue, mean, variance, skewness, kurtosis</i>	80%	21
0,5	<i>mean red, mean green, mean blue, mean, variance, kurtosis</i>	78%	39
0,7	<i>mean red, mean green, mean, variance, kurtosis</i>	88%	32
1	<i>mean red, mean green, mean, variance</i>	50%	1000

4. KESIMPULAN

Dari hasil analisa yang dilakukan berdasarkan percobaan yang dilakukan, dapat disimpulkan beberapa hal sebagai berikut:

1. Nilai *information gain* untuk setiap fitur sangat dipengaruhi oleh nilai data yang ada didalamnya. Semakin banyak variasi nilai yang ada maka semakin tinggi kemungkinan fitur memiliki nilai *information gain* yang tinggi. Namun semakin banyak variasi nilai yang ada pada sebuah fitur memungkinkan fitur tersebut memberikan hasil yang kurang baik jika digunakan dalam proses *clustering*.
2. Seleksi fitur *information gain* sangat berpengaruh terhadap proses dan hasil *clustering* menggunakan algoritma *fuzzy c-means*. Akurasi yang diperoleh jika tidak menggunakan seleksi fitur adalah 50%, sedangkan akurasi yang diperoleh jika dilakukan menggunakan seleksi fitur *information gain* dengan nilai threshold 0,7 adalah 88%.

DAFTAR PUSTAKA

- AAK. 1981. *Petunjuk Bercocok Tanam Cengkeh*. Yogyakarta : Kanisius.
- Ahmad, Usman. 2005. *Pengolahan Citra Digital dan Teknik Pemrogramannya*. Yogyakarta : Graha Ilmu.
- Hidayatsyah, M R. 2013. *Penerapan Metode Decision Tree dalam Pemberian Pinjaman Kepada Debitur dengan Algoritma C4.5*. Tugas Akhir :UIN Sultan Syarif Kasim Riau.
- Kusumadewi, S & Purnomo, H. 2013. *Aplikasi Logika Fuzzy Untuk Pendukung Keputusan*. Yogyakarta : Graha Ilmu.
- Maulana, M.R & Chasanah, T.T. 2016. *Peningkatan performa algoritma C4.5 dengan seleksi fitur information gain untuk klasifikasi persetujuan kredit*. IC-Tech Volume XI (2).
- Maulana, M.R& Karomi, M.A. 2015. *Information Gain untuk Mengetahui Pengaruh Atribut Terhadap Klasifikasi Persetujuan Kredit*. Jurnal LITBANG Kota Pekalongan Volume 9.
- Munir, R. 2004. *Pengolahan Citra Digital dengan Pendekatan Algoritmik*. Bandung : Informatika.
- Wahyuni, E.S.2016. *Penerapan Metode Seleksi Fitur untuk Meningkatkan Hasil Diagnosis Kanker*. Jurnal SIMETRIS Vol 7 (1), ISSN: 2252-4983.
- Sari, B.N. 2016. *Implementasi Teknik Seleksi Fitur Information Gain pada Algoritma Klasifikasi Machine Learning untuk Prediksi Performa Akademik Siswa*. Seminar Nasional Teknologi Informasi dan Multimedia, ISSN: 2302-3805.
- Wang, L. 1997. *A Course in Fuzzy Systems and Control*. Prentice-Hall International, Inc.
- Yuan, B& Klir, G J. 1995. *Fuzzy Sets and Fuzzy Logic: Theori andApplications*. Prentice-Hall International, Inc.
-

LAMPIRAN(Hasil Ekstraksi Fitur)

No	Citra	R	G	B	Mean	Var	Skew	Kurt	Ent
1	TS_1.JPG	106	104	105	105	1234	0	-1	-7
2	TS_2.JPG	72	70	70	71	814	0	0	-7
3	TS_3.JPG	123	122	124	123	1826	0	-1	-7
4	TS_4.JPG	68	66	65	66	998	0	0	-7
5	TS_5.JPG	114	111	111	112	1448	-1	0	-7
6	TS_6.JPG	105	99	94	99	1716	0	-1	-7
7	TS_7.JPG	89	83	76	82	2165	0	-1	-7
8	TS_8.JPG	97	93	86	92	1047	0	-1	-7
9	TS_9.JPG	93	90	84	89	1053	0	-1	-7
10	TS_10.JPG	108	102	96	102	1346	0	-1	-7
11	TS_11.JPG	114	109	100	108	1475	0	-1	-7
12	TS_12.JPG	95	91	83	90	2091	0	-1	-7
13	TS_13.JPG	109	105	98	104	2246	0	-1	-7
14	TS_14.JPG	90	83	72	82	1175	0	-1	-7
15	TS_15.JPG	106	93	74	91	1730	0	-1	-7
16	TS_16.JPG	66	61	56	61	1875	1	0	-7
17	TS_17.JPG	75	71	70	72	947	0	-1	-7
18	TS_18.JPG	88	82	75	82	2166	0	-1	-7
19	TS_19.JPG	94	89	82	88	1235	0	-1	-7
20	TS_20.JPG	110	103	96	103	1523	0	-1	-7
21	TS_21.JPG	135	126	115	125	1139	-1	0	-7
22	TS_22.JPG	86	82	75	81	2075	0	-1	-7
23	TS_23.JPG	108	106	108	107	1226	0	-1	-7
24	TS_24.JPG	68	66	66	67	1037	0	-1	-7
25	TS_25.JPG	111	105	99	105	1678	0	-1	-7
26	TTS_1.JPG	176	142	108	142	406	1	2	-6
27	TTS_2.JPG	172	132	86	130	199	0	12	-5
28	TTS_3.JPG	153	134	113	133	1634	0	0	-7
29	TTS_4.JPG	136	123	106	122	1430	0	0	-7
30	TTS_5.JPG	180	163	142	162	712	0	0	-7
31	TTS_6.JPG	175	143	123	147	503	-1	2	-6
32	TTS_7.JPG	168	124	91	128	307	-1	2	-6
33	TTS_8.JPG	164	128	98	130	889	-1	1	-7
34	TTS_9.JPG	139	119	88	115	1004	-1	2	-7
35	TTS_10.JPG	154	129	96	126	680	-1	2	-6
36	TTS_11.JPG	208	158	80	149	1014	-1	0	-7

No	Citra	R	G	B	Mean	Var	Skew	Kurt	Ent
37	TTS_12.JPG	203	159	84	149	847	-1	1	-7
38	TTS_13.JPG	218	171	98	162	1098	-1	0	-7
39	TTS_14.JPG	176	118	51	115	1015	0	0	-7
40	TTS_15.JPG	218	162	83	154	745	-1	1	-7
41	TTS_16.JPG	191	139	68	133	1923	-1	0	-7
42	TTS_17.JPG	202	149	71	141	1477	-1	0	-7
43	TTS_18.JPG	171	142	89	134	1223	0	0	-7
44	TTS_19.JPG	180	140	88	136	443	-1	1	-6
45	TTS_20.JPG	171	143	103	139	614	0	1	-7
46	TTS_21.JPG	169	140	98	136	885	-1	1	-7
47	TTS_22.JPG	178	143	96	139	777	-1	1	-7
48	TTS_23.JPG	195	161	124	160	812	-1	2	-7
49	TTS_24.JPG	187	153	121	154	803	-1	2	-7
50	TTS_25.JPG	182	134	87	134	478	-1	2	-6
