

**STUDI PERBANDINGAN KLASIFIKASI MULTISPEKTRAL
MAXIMUM LIKELIHOOD DAN SUPPORT VECTOR MACHINE
UNTUK PEMETAAN PENUTUP LAHAN**

Ardhian Nur Rahman Wal Hamdir
ardhian_nr@yahoo.com

Sigit Herumurti B.S.
sigit@geo.ugm.ac.id

Abstract

There are several methods in remote sensing developed to obtain landcover information, one of those is maximum likelihood classification, which is a classification method that is considered to be the most statistically stable to obtain landcover informations derived from remote sensing image. This classification has the assumption that remote sensing data inputs were normally distributed. To anticipate the remote sensing data that were not normally distributed, the approach is to use a classification based on the non-parametrical statistics such as support vector machine classification. The purpose of this research is to obtain the accuracy levels of both maximum likelihood and support vector machine classification based on their different approaches.

Accuracy assesment results using Confusion Matrix and Kappa Coefficient Index showing that Support Vector Machine classification using Linear Kernel gives 65,88% accuracy which is the highest overall accuracy within the other methods and 0,6406 Kappa Coefficient Index.

Keyword: multispectral classification, landcover, maximum likelihood, support vector machine

Abstrak

Dalam penginderaan jauh terdapat beberapa metode yang dikembangkan untuk mendapatkan informasi mengenai penutup lahan, salah satunya yaitu klasifikasi maximum likelihood, yang merupakan metode klasifikasi yang dianggap paling mapan secara statistik untuk memetakan informasi penutup lahan dari citra penginderaan jauh. Klasifikasi ini memiliki asumsi bahwa data penginderaan jauh masukan yang digunakan harus terdistribusi normal. Untuk mengantisipasi adanya data penginderaan jauh yang tidak terdistribusi normal, pendekatan yang digunakan adalah dengan menggunakan klasifikasi yang didasarkan pada logika statistik non-parametrik seperti Klasifikasi Support Vector Machine. Perbedaan pendekatan inilah yang selanjutnya menjadi tujuan dari penelitian ini, yaitu mengetahui tingkat akurasi antara Klasifikasi Maximum Likelihood dan Klasifikasi Support Vector Machine.

Hasil perhitungan menggunakan Confusion Matrix dan Analisis Koefisien Kappa menunjukkan bahwa Klasifikasi Support Vector Machine (SVM) dengan Kernel Linear memiliki nilai akurasi paling tinggi dengan nilai akurasi keseluruhan sebesar 65,88% dan nilai Indeks Koefisien Kappa sebesar 0,6046.

Kata kunci: klasifikasi multispektral, penutup lahan, maximum likelihood, support vector machine

PENDAHULUAN

Informasi yang diturunkan dari data penginderaan jauh menjadi semakin penting untuk kegiatan perencanaan yang berkaitan dengan lingkungan baik pada skala lokal, regional maupun global (Johannes dkk, 2003 dalam Jensen 2005). Informasi yang diturunkan dari data penginderaan jauh ini dapat berupa peta tematik atau data statistik. Informasi ini harus akurat karena keputusan – keputusan penting di seluruh dunia akan dibuat berdasarkan informasi tersebut (Muchoney and Strahler, 2002; Kyriakidis dkk, 2003 dalam Jensen 2005). Salah satu informasi penting yang dapat diturunkan dari data penginderaan jauh adalah informasi mengenai penutup lahan dan penggunaan lahan.

Dalam penginderaan jauh terdapat beberapa metode yang dikembangkan untuk mendapatkan informasi mengenai penutup lahan. Salah satu metode yang banyak digunakan adalah klasifikasi multispektral secara digital berdasarkan analisis statistik. Klasifikasi penutup lahan atau penggunaan lahan berdasarkan teknik pengenalan pola secara statistik yang diaplikasikan pada data penginderaan jauh multispektral adalah salah satu dari beberapa metode yang paling banyak digunakan dalam proses pengumpulan informasi (Narumalani dkk, 2002 dalam Jensen, 2005).

Klasifikasi *maximum likelihood* merupakan salah satu metode klasifikasi berdasarkan algoritma statistik parametrik dimana terdapat asumsi bahwa data penginderaan jauh terdistribusi normal dan bentuk dari kurva kepadatan dapat diketahui (Duda dkk, 2001 dalam Jensen, 2005). Klasifikasi ini juga merupakan algoritma yang secara statistik dianggap paling mapan (Danoedoro, 1996), sehingga sering digunakan dalam memetakan informasi penutup lahan dari citra penginderaan jauh. Meskipun dianggap paling mapan, klasifikasi ini masih berhadapan dengan masalah, yaitu klasifikasi ini memiliki asumsi bahwa data penginderaan jauh masukan yang digunakan harus terdistribusi normal. Untuk mengantisipasi adanya data penginderaan jauh yang tidak

terdistribusi normal, salah satu pendekatan yang digunakan adalah dengan menggunakan klasifikasi yang didasarkan pada logika statistik non-parametrik.

Berbeda dengan klasifikasi *support vector machine* yang termasuk ke dalam analisis statistik non parametrik. Klasifikasi *support vector machine* tidak memiliki asumsi bahwa *training data* yang dipakai harus terdistribusi normal karena klasifikasi ini mengidentifikasi kelas spektral dan menganalisis tingkat hubungannya dengan masing – masing piksel. Klasifikasi ini juga mampu menyediakan hasil yang relatif memuaskan untuk data penginderaan jauh yang kompleks dan terdapat banyak gangguan (ENVI *User's Guide*, 2008).

TUJUAN PENELITIAN

1. Mengetahui nilai akurasi hasil klasifikasi *maximum likelihood* dalam mendapatkan informasi penutup lahan.
2. Mengetahui nilai akurasi hasil klasifikasi *support vector machine* dalam mendapatkan informasi penutup lahan.
3. Mengetahui metode yang paling baik untuk mendapatkan informasi penutup lahan berdasarkan nilai akurasi antara klasifikasi *maximum likelihood* dan klasifikasi *support vector machine*.

METODE PENELITIAN

Data yang digunakan pada penelitian ini adalah :

1. Citra EO-1 ALI tahun 2009 sebagian kabupaten Bantul.
2. Peta Rupa Bumi Indonesia skala 1:25.000 daerah kabupaten Bantul.

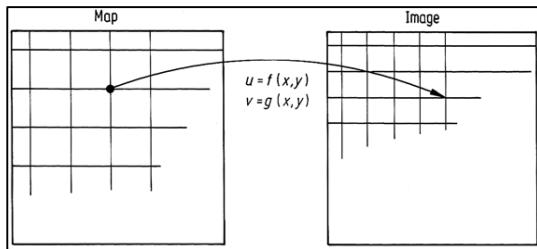
Alat yang digunakan pada penelitian ini adalah:

1. Seperangkat komputer.
2. Perangkat lunak ENVI 4.5 untuk pengolahan citra secara digital.
3. GPS untuk navigasi pada saat cek lapangan.

Koreksi Geometrik

Tahapan koreksi geometrik yang dilakukan adalah memasukkan beberapa *ground control point* (GCP) dari suatu peta ataupun citra lain yang sudah terkoreksi dengan datum lokal. Terdapat beberapa metode *resampling* yang dapat dilakukan, namun untuk kepentingan klasifikasi dilakukan koreksi geometri dengan orde 1 (*affine*)

yaitu *nearest neighbour resampling* karena metode resampling ini hanya memindahkan titik tengah dari suatu piksel ke dalam titik terdekat dari suatu lokasi pada peta atau citra rujukan, kemudian piksel tersebut ditransformasi ke dalam koordinat x,y yang baru tanpa merubah nilai piksel aslinya seperti ditunjukkan pada gambar 3.1



Gambar 3.1. Penggunaan *nearest neighbour resampling* untuk mentransformasikan piksel ke dalam koordinat yang baru. Sumber : Richard, 2005.

Layer Stacking

Layer Stacking digunakan untuk menggabungkan citra dari beberapa saluran dengan resolusi spasial yang berbeda menjadi satu layer yang sama, misalnya file saluran pankromatik pada ALI dengan resolusi spasial 10m digabungkan dengan file saluran multispektral dengan resolusi 30m, menjadi 1 file tunggal dengan resolusi spasial 30m, namun mengandung informasi nilai spektral dari kedua saluran tersebut.

Pengambilan Training Area (Daerah Contoh)

Training Area (Daerah Contoh) merupakan objek-objek klas penutup lahan yang menjadi masukan dalam klasifikasi *maximum likelihood* dan *support vector machine*. Selain itu, sampel yang dipilih harus memiliki nilai separabilitas yang tinggi antara masing – masing klas penutup lahan, yaitu mendekati 2. Nilai separabilitas ini didapat dari algoritma *transformed divergence* yang digunakan untuk menghitung nilai indeks separabilitas.

Klasifikasi *Maximum Likelihood* (MLC)

Klasifikasi *Maximum Likelihood* (MLC) menggunakan asumsi bahwa data statistik dari masing-masing kelas pada masing-masing saluran terdistribusi normal dan menhitung probabilitas suatu piksel untuk masuk dalam kelas tertentu. Ambang batas probabilitas nilai piksel (*threshold*) ditentukan dalam proses klasifikasi. apabila nilai probabilitas tertinggi dari suatu piksel lebih rendah dari nilai ambang batas yang ditetukan, maka piksel tersebut masuk dalam kelas tak terkласifikasi (*unclassified pixel*). Algoritma

klasifikasi *Maximum Likelihood* dalam Danoedoro (1996) ditunjukkan pada persamaan berikut ini :

$$D = \ln(a_c) - [0.5 \ln(|\text{Cov}_c|) - 0.5(\mathbf{X} - \mathbf{M}_c)^T (\text{Cov}_c^{-1}) (\mathbf{X} - \mathbf{M}_c)]$$

dimana :

- D : jarak yang diberi bobot
- c : suatu kelas tertentu
- \mathbf{X} : vektor piksel yang diklasifikasi
- \mathbf{M}_c : vektor rerata sampel kelas c
- a_c : presentase probabilitas sembarang piksel untuk menjadi kelas c
- Cov_c : matriks kovariansi piksel – piksel pada sampel kelas c
- $|\text{Cov}_c|$: determinan Cov_c
- Cov_c^{-1} : inversi Cov_c
- \ln : logaritma natural
- T : fungsi transposisi (aljabar matriks)

Klasifikasi *Support Vector Machine* (SVM)

Metode yang diterapkan dalam SVM untuk mendapatkan informasi mengenai suatu objek, adalah dengan memisahkan objek yang satu dengan objek yang lainnya menggunakan suatu *kernel* yang bekerja berdasarkan jarak euklidian pada *feature space* multi - dimensi (Kohram, 2008). Beberapa fungsi *kernel* yang dapat diaplikasikan untuk data penginderaan jauh adalah linear, polynomial, *radial basis function* (RBF) , dan *sigmoid* (Hsu dkk, 2007). Fungsi tersebut secara matematis direpresentasikan dalam persamaan berikut ini:

Linear	$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$
Polynomial	$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{g} \mathbf{x}_i^T \mathbf{x}_j + r) d, g > 0$
RBF	$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-g \ \mathbf{x}_i - \mathbf{x}_j\ ^2), g > 0$
Sigmoid	$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\mathbf{g} \mathbf{x}_i^T \mathbf{x}_j + r)$

Pengambilan Sampel Lapangan

Pengambilan sampel lapangan mengacu pada skema *stratified random sampling*, dimana jumlah minimum sampel dipilih dari masing-masing strata (klas penutup lahan). Ukuran sampel lapangan didasarkan pada teori probabilitas binomial menurut Fitzpatrick-Lins(1981) yaitu :

$$N = Z^2 (p)(q)/E^2$$

dimana p adalah persentase akurasi yang diharapkan dari keseluruhan wilayah, $q = 100 - p$, E adalah kesalahan yang diperbolehkan, dan $Z= 2$ dari standar deviasi untuk tingkat kepercayaan 95%, untuk akurasi yang diharapkan sebesar 85% dan kesalahan yang dapat diterima sebesar 10%

maka ukuran sampel yang dibutuhkan minimum sebesar 51 sampel.

HASIL DAN PEMBAHASAN

Hasil Klasifikasi *Maximum Likelihood* (MLC) dan *Support Vector Machine* (SVM)

Hasil dari proses Klasifikasi *Maximum Likelihood* berupa citra klasifikasi dengan 10 kelas penutup lahan dengan nilai akurasi keseluruhan sebesar 54,66% dan nilai koefisien kappa sebesar 0,4901. Nilai akurasi pembuat peta (producer accuracy) paling tinggi ditunjukkan oleh kelas vegetasi berkayu kerapatan tinggi (C213) yaitu sebesar 80,95 % dan nilai akurasi pengguna peta (user accuracy) paling tinggi ditunjukkan oleh kelas penutup lahan atap tanah liat (C421) yaitu sebesar 94,59%. Hal ini menunjukkan bahwa kecenderungan kelas atap tanah liat merepresentasikan objek lain di lapangan dan kecenderungan pembuat peta menganggap kelas lain pada saat proses klasifikasi menjadi kelas penutup lahan atap tanah liat sangat rendah. Sedangkan nilai akurasi pembuat peta (producer accuracy) paling rendah ditunjukkan oleh kelas penutup lahan vegetasi tak berkayu kerapatan sedang (C222) yaitu sebesar 1,39 % dan nilai akurasi pengguna (user accuracy) paling rendah juga ditunjukkan oleh kelas penutup lahan vegetasi tak berkayu kerapatan sedang (C222) yaitu sebesar 11,11%.

Hasil dari proses Klasifikasi *Support Vector Machine* (SVM) dengan kernel Linier berupa citra klasifikasi dengan 10 kelas penutup lahan dengan nilai akurasi keseluruhan sebesar 57,98 % dan nilai koefisien kappa sebesar 0,5297. Nilai akurasi pembuat peta (producer accuracy) paling tinggi ditunjukkan oleh kelas atap seng (C431) yaitu sebesar 88,89 % dan nilai akurasi pengguna peta (user accuracy) paling tinggi ditunjukkan oleh kelas penutup lahan atap tanah liat (C421) yaitu sebesar 93,65%. Sedangkan nilai akurasi pembuat peta (producer accuracy) paling rendah ditunjukkan oleh kelas penutup lahan vegetasi tak berkayu kerapatan sedang (C222) yaitu sebesar 27,78 % dan nilai akurasi pengguna (user accuracy) paling rendah ditunjukkan oleh kelas tubuh air (C12) yaitu sebesar 17,11%.

Hasil dari proses Klasifikasi *Support Vector Machine* (SVM) dengan kernel Polinomial berupa citra klasifikasi dengan 10 kelas penutup lahan dengan nilai akurasi keseluruhan sebesar 55,92 % dan nilai koefisien kappa sebesar 0,5064. Nilai akurasi pembuat peta (producer accuracy) paling tinggi ditunjukkan oleh kelas atap seng (C431) yaitu sebesar 86,11 % dan nilai akurasi pengguna

peta (user accuracy) paling tinggi ditunjukkan oleh kelas penutup lahan atap tanah liat (C421) yaitu sebesar 91,94%. Sedangkan nilai akurasi pembuat peta (producer accuracy) paling rendah ditunjukkan oleh kelas penutup lahan vegetasi tak berkayu kerapatan sedang (C222) yaitu sebesar 25 % dan nilai akurasi pengguna (user accuracy) paling rendah ditunjukkan oleh kelas tubuh air (C12) yaitu sebesar 18,18%.

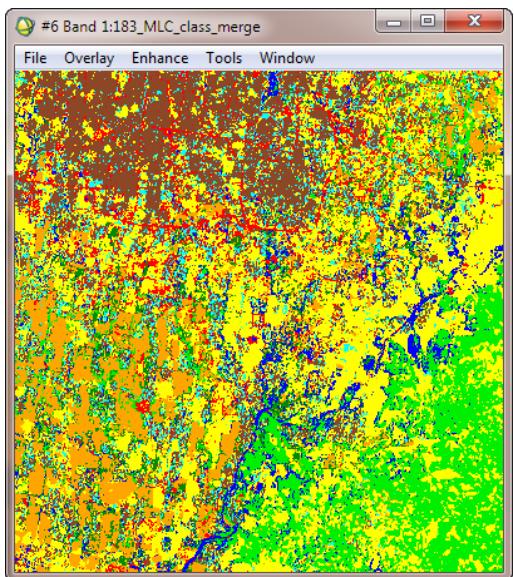
Hasil dari proses Klasifikasi *Support Vector Machine* (SVM) dengan kernel RBF berupa citra klasifikasi dengan 10 kelas penutup lahan dengan nilai akurasi keseluruhan sebesar 56,24 % dan nilai koefisien kappa sebesar 0,5099. Nilai akurasi pembuat peta (producer accuracy) paling tinggi ditunjukkan oleh kelas atap seng (C431) yaitu sebesar 86,11 % dan nilai akurasi pengguna peta (user accuracy) paling tinggi ditunjukkan oleh kelas penutup lahan atap tanah liat (C421) yaitu sebesar 91,94%. Sedangkan nilai akurasi pembuat peta (producer accuracy) paling rendah ditunjukkan oleh kelas penutup lahan vegetasi tak berkayu kerapatan sedang (C222) yaitu sebesar 26,39 % dan nilai akurasi pengguna (user accuracy) paling rendah ditunjukkan oleh kelas tubuh air (C12) yaitu sebesar 18,18%.

Hasil dari proses Klasifikasi *Support Vector Machine* (SVM) dengan kernel RBF berupa citra klasifikasi dengan 10 kelas penutup lahan dengan nilai akurasi keseluruhan sebesar 54,34 % dan nilai koefisien kappa sebesar 0,4886. Nilai akurasi pembuat peta (producer accuracy) paling tinggi ditunjukkan oleh kelas atap seng (C431) yaitu sebesar 86,11 % dan nilai akurasi pengguna peta (user accuracy) paling tinggi ditunjukkan oleh kelas penutup lahan atap tanah liat (C421) yaitu sebesar 92,86%. Sedangkan nilai akurasi pembuat peta (producer accuracy) paling rendah ditunjukkan oleh kelas penutup lahan vegetasi tak berkayu kerapatan sedang (C222) yaitu sebesar 23,61 % dan nilai akurasi pengguna (user accuracy) paling rendah ditunjukkan oleh kelas tubuh air (C12) yaitu sebesar 19,23%.

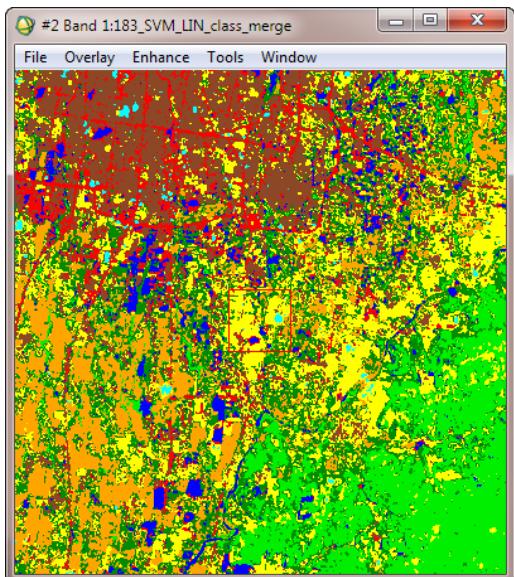
Penilaian Akurasi Hasil Klasifikasi

Hasil uji akurasi dilakukan dengan menggunakan metode *Confusion Matrix* 633 piksel dari Citra hasil klasifikasi yang diambil secara independen dengan bantuan interpretasi visual citra resolusi tinggi Quickbird. Interpretasi visual dilakukan dengan menentukan matriks 200x200 piksel citra Quickbird yang dianggap merepresentasikan kondisi matriks 4x4 piksel pada citra hasil klasifikasi. Hasil perhitungan menggunakan *Confusion Matrix* dan Analisis

Koefisien Kappa seperti pada Tabel 1 dan Tabel 2 menunjukkan bahwa Klasifikasi *Support Vector Machine* (SVM) dengan *Kernel Linear* memiliki nilai akurasi paling tinggi dari 4 metode lainnya baik untuk 10 kelas penutup lahan maupun 8 kelas penutup lahan.



Gambar 1. Citra Hasil Klasifikasi Multispektral *Maximum Likelihood* (MLC) 8 kelas penutup lahan. Sumber : Pengolahan Citra, 2014



Gambar 2. Citra Hasil Klasifikasi Multispektral *Support Vector Machine* (SVM) *Kernel Linear* 8 kelas penutup lahan. Sumber : Pengolahan Citra, 2014

KESIMPULAN

1. Klasifikasi *Maximum Likelihood* (MLC) memiliki nilai akurasi paling rendah kedua setelah Klasifikasi *Support Vector Machine*

dengan *Kernel Sigmoid* dengan Nilai Akurasi Keseluruhan sebesar 60,35% dan Nilai Koefisien Kappa sebesar 0,5424, sedangkan Klasifikasi *Support Vector Machine* (SVM) dengan *Kernel Linear* memiliki nilai akurasi paling tinggi dibandingkan dengan klasifikasi *Maximum Likelihood* (MLC) dengan Nilai Akurasi Keseluruhan sebesar 65,88% dan Nilai Koefisien Kappa sebesar 0,6046

2. Klasifikasi *Support Vector Machine* (SVM) memperlihatkan kemampuan yang lebih baik dalam menurunkan informasi penutup lahan menggunakan data penginderaan jauh dengan tingkat gangguan (*noise*) yang relatif besar dan tanpa memperhatikan normalitas data penginderaan jauh masukan.

DAFTAR PUSTAKA

- BPS Kabupaten Bantul. (2014). *Bantul dalam Angka* 2014. <http://bantulkab.bps.go.id/index.php?id=kda2014>. diakses 18 Oktober 2014.
- BPS Kota Yogyakarta. (2014). *Yogyakarta dalam Angka* 2014. http://jogjakota.bps.go.id/index.php?hal=publikasi_detil&id=13. diakses 18 Oktober 2014.
- Chander, G., Markham, B. L., Helder, D. L. (2009). Summary of Current Radiometric Calibration Coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI Sensors. *Remote Sensing of Environment*, 113, 893-903.
- Danoedoro, P. (1996). Pengolahan Citra Digital: Teori dan Aplikasinya dalam Bidang Penginderaan Jauh. Yogyakarta: Fakultas Geografi UGM.
- Danoedoro, P., dan Phinn, S.R. (2005). Detailed Land Cover Mapping by Introducing Higher-Spatial Resolution Panchromatic Bands in Multi-Spectral Classification: Examples Using Landsat ETM+ and Quickbird Imagery. *MapAsia 2005 Conference: Empowering People Through Geospatial Information, GIS Development*: Jakarta.
- Foody, M. G., dan Mathur, A. (2004a). A Relative Evaluation of Multiclass Image Classification by Support Vector Machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 1335-1343.
- Foody, M. G., dan Mathur, A. (2004b). Toward Intelligent Training of Supervised Image

- Classifications: Directing Training Data Acquisition for SVM Classification. *Remote Sensing of Environment*, 93, 107-117.
- Gidudu, G. Hulley dan T. Marwala. (2007). Image classification using SVMs: One-against-one vs One-against-all. *Proceeding of the 28th Asian Conference on Remote Sensing*, 2, 801-806.
- Hsu, C.W., Chan, C.C., dan Lin, C.J. (2003). *A Practical Guide to Support Vector Classification*. National Taiwan University: Taiwan.
- Janssen, L. L. F. (2000). *Principles of Remote Sensing : An Introductory Textbook*. Enschede: ITC
- Kohram, Mojtaba. (2008). Composite Kernels for Support Vector Classification of Hyperspectral Data. *MICAI 2008: Advances in Artificial Intelligence. Lecture Notes in Computer Science*, 5317, 360-370
- Lillesand, T. dan Kiefer, Ralph W. (1990). *Penginderaan Jauh dan Interpretasi Citra*. Edisi Terjemahan. Yogyakarta: Gadjah Mada University Press.
- Malingreau, J. P. dan R. Christiani. (1982). *Land Cover/Land Use Classification for Indonesia*. Yogyakarta: PUSPICS UGM.
- Oommen, T., Misra, D, et al. (2008). An Objective Analysis of Support Vector Machine Based Classification for Remote Sensing. *Math Geosci*. Nomor 40, hal 409-424.
- Richards, J.A dan Jia, X. (2005), *Remote Sensing Digital Image Analysis*, Springer-Verlag, Berlin
- Sutanto. (1986). Penginderaan Jauh. Jilid 1. Yogyakarta: Gadjah Mada University Press.
- Vapnik, V.N. (1995). *The Nature of Statistical Learning Theory*. Springer: New York. Hal 188.

Tabel 1. Penilaian Akurasi Klasifikasi *Maximum Likelihood* (MLC) dan *Support Vector Machine* (SVM) untuk 10 kelas penutup lahan.

Nilai Akurasi	Metode	MLC	SVM Linier	SVM Polinomial	SVM RBF	SVM Sigmoid
Akurasi Keseluruhan (%)		54,57	57,98	55,92	56,24	54,34
Koefisien Kappa		0,4901	0,5297	0,5064	0,5099	0,4886

Sumber : Pengolahan Citra, 2014

Tabel 2. Penilaian Akurasi Klasifikasi *Maximum Likelihood* (MLC) dan *Support Vector Machine* (SVM) untuk 8 kelas penutup lahan.

Nilai Akurasi	Metode	MLC	SVM Linier	SVM Polinomial	SVM RBF	SVM Sigmoid
Akurasi Keseluruhan (%)		60,35	65,88	64,14	64,30	63,51
Koefisien Kappa		0,5424	0,6046	0,5840	0,5857	0,5761

Sumber : Pengolahan Citra, 2014