

METODE CTA DENGAN TEKNIK DATA MINING CITRA LANDSAT-8 UNTUK KLASIFIKASI PENGGUNAAN LAHAN

Cahya Budi Perwitagama¹, M. Pramono Hadi², dan Nur Mohammad Farda³

Fakultas Geografi, Universitas Gadjah Mada, Yogyakarta, Indonesia^{1,2,3}
cahyabpgeo@gmail.com

Diterima : Januari 2015 ; Direvisi : Maret 2014.; Dipublikasikan: 30 September 2015

ABSTRAK Kajian mengenai metode non-parametrik *Classification Tree Analysis* (CTA) menggunakan teknik *data mining* untuk aplikasi penginderaan jauh masih belum banyak dilakukan. Sehingga diperlukan penelitian mengenai kemampuan CTA dalam menangani data yang cukup banyak, dengan memanfaatkan kelebihan CTA untuk aplikasi penginderaan jauh. Kombinasi parameter CTA dan data masukan, serta penerapannya pada dua skema klasifikasi yang berbeda tingkat kerinciannya, memerlukan pengujian terkait dengan tingkat akurasi yang dihasilkan. Penelitian ini bertujuan melakukan simulasi dari beberapa kombinasi parameter untuk mengetahui tingkat akurasi hasil klasifikasi, dan memperoleh pohon keputusan dari hasil KDD. Serta menganalisis akurasi metode non-parametrik CTA dengan teknik *data mining* untuk klasifikasi penggunaan lahan menggunakan citra Landsat-8 OLI, dan menerapkan hasil KDD pada daerah lain. Klasifikasi diperoleh dengan melakukan simulasi beberapa parameter CTA dan data masukan. Parameter aturan pemisah (*splitting rules*) dalam CTA, yaitu *Ratio*, *Entropy*, dan *Gini*. Parameter pemangkas (*pruning*), yaitu 0%, 1%, 5%, dan 10%. Beberapa masukan data klasifikasi, antara lain adalah citra Landsat-8 tujuh saluran, transformasi citra (NDVI, NDWI, BI, NDBI, dan PCA), serta filter tekstur *variance* dan *mean* (jendela bergerak 3x3 dan 5x5). Data non-spektral, yaitu data ketinggian dan data kemiringan lereng. Dua tingkat skema klasifikasi penggunaan lahan, yaitu Level I (5 kelas) dan Level II (8 kelas). Pohon keputusan yang diperoleh dari hasil pembelajaran dengan akurasi terbaik kemudian diterapkan pada daerah lain yang memiliki karakteristik mirip. Hasil penelitian menunjukkan bahwa kombinasi parameter terbaik adalah atribut pemisah *Gini*, *pruning* 1%, filter tekstur dengan jendela bergerak 5x5, dan skema Level I yaitu dengan akurasi keseluruhan 96,71%, *kappa* 0,9504, dan waktu proses 3,388 detik. Penerapannya pada daerah lain, menghasilkan akurasi keseluruhan 93,27% dengan *kappa* 0,8923. Tingkat akurasi terbaik yang diperoleh pada daerah penelitian maupun penerapannya pada daerah lain, lebih besar dari 90%. Sehingga diharapkan metode ini dapat menjadi alternatif metode untuk terapan kebijakan penggunaan lahan, dan klasifikasi penggunaan lahan berbasis penutup lahan setara dengan skala 1:100.000.

Kata kunci: *Classification Tree Analysis* (CTA); *data mining*; klasifikasi penggunaan lahan; Landsat-8.

ABSTRACT The study of non-parametric methods *Classification Tree Analysis* (CTA) using *data mining* techniques for remote sensing applications is still much to do, so it requires studies on the CTA's ability to handle quite a lot of data, by utilizing CTA advantages for remote sensing applications. The combination of parameter CTA and data input, as well as its application on two different detail levels of land use classification scheme, require testing related to the level of accuracy that is generated. This study aimed to simulate multiple combinations of parameters to determine the level of accuracy of the classification results, and obtain a decision tree from KDD results. And to analyze the accuracy of non-parametric methods CTA with *data mining* techniques for land use classification using Landsat-8 OLI, and apply the results of KDD on another area. The classification is obtained by performing simulations on several parameters CTA and data input. There are three *splitting rules* parameter in CTA, i.e. *Ratio*, *Entropy*, and *Gini*. *Pruning* parameter, i.e. 0%, 1%, 5%, and 10%. There are several inputs of the classification data, namely seven bands of Landsat-8 OLI imagery, image transformation (NDVI, NDWI, BI, NDBI, and PCA), as well as texture filter *variance* and texture filter *mean* (moving window 3x3 and 5x5). Non-spectral data, i.e. elevation data and slope data. And two-level land use classification scheme, i.e. Level I (5 classes) and Level II (8 classes). The decision tree that obtained from the best accuracy of the learning outcomes then applied to another area with similar characteristics. The results showed that the best parameter combination are a *splitting attribute Gini*, *pruning* 1%, texture filter with 5x5 moving window, and Level I scheme, with an accuracy of 96.71%, *kappa* 0.9504, and processing time 3.388 sec. Its application on another area, resulting an overall accuracy 93.27% with *kappa* 0.8923. The best accuracy rate obtained in this study and its application on another area was greater than 90%. Therefore this method is expected could be an alternative method for land use policy application, and land use classification based on land cover commensurate to a scale of 1:100,000.

Key words: *Classification Tree Analysis* (CTA); *data mining*; land use classification; Landsat-8.

PENDAHULUAN

Kajian mengenai metode non-parametrik *Classification Tree Analysis* (CTA) menggunakan teknik *data mining* untuk aplikasi penginderaan jauh masih belum banyak dilakukan,

sehingga diperlukan penelitian mengenai kemampuan CTA dalam menangani data yang cukup banyak, dengan memanfaatkan kelebihan CTA untuk aplikasi penginderaan jauh. Kelebihan CTA untuk aplikasi penginderaan jauh, antara lain CTA merupakan metode

non-parametrik, sederhana, mampu menangani hubungan non-linear dan *noise* antara fitur-fitur masukan dan label kelas, dan efisiensi komputasi.

Penggunaan CTA ini dilakukan dengan caramemberikan masukan dari beberapa saluran citra Landsat-8 OLI, transformasi, pemfilteran serta data tambahan, yakni data ketinggian dan kemiringan lereng. Citra Landsat-8 merupakan citra satelit dengan cakupan *scene* yang luas dan memiliki resolusi spasial 30 meter yang dapat digunakan untuk kajian dengan skala regional, setara dengan wilayah kabupaten. Landsat-8 memiliki jumlah saluran yang cukup banyak, datanya baru sehingga memiliki informasi terkini, dan dapat diakses dengan mudah. Kemampuan citra Landsat-8 tersebut akan sangat membantu dalam pemetaan penggunaan lahan di Indonesia yang memiliki wilayah yang luas. Hal ini selanjutnya dapat dimanfaatkan untuk mempercepat pembangunan Indonesia.

Informasi sebaran penggunaan lahan diperlukan oleh pemerintah, swasta, maupun masyarakat. Informasi ini dapat digunakan untuk mengetahui penggunaan lahan pada masa lalu, saat ini, prediksi di masa depan, maupun untuk perencanaan dan penyusunan kebijakan. Skema klasifikasi penggunaan lahan yang digunakan dalam penelitian akan mengacu pada skema klasifikasi penggunaan lahan yang dilakukan BPN. Skema klasifikasi tersebut akan dimodifikasi agar sesuai dengan informasi yang diperoleh menggunakan metode penginderaan jauh. Perkembangan teknologi penginderaan jauh ini diiringi dengan perkembangan pengolahan citra digital. Semakin meningkatnya kemampuan komputer dalam mengolah data yang cukup besar, sangat membantu dalam pengembangan berbagai teknik dan analisis citra digital, salah satunya adalah dapat dimanfaatkan untuk klasifikasi.

Klasifikasi multispektral pada umumnya menghasilkan peta penutup lahan. Penutup lahan dapat diinterpretasi menggunakan citra penginderaan jauh. Dari data dan informasi tersebut dapat diturunkan menjadi informasi-informasi lain, seperti penggunaan lahan. Dalam algoritma klasifikasi multispektral pada umumnya dibagi menjadi dua, yaitu algoritma parametrik dan algoritma non-parametrik. Algoritma parametrik mengasumsikan bahwa kelas terdistribusi normal dan memerlukan perkiraan distribusi parameter, seperti *mean vector* dan *covariance matrix* dalam melakukan klasifikasi, contoh algoritma parametrik antara lain adalah *maximum likelihood* dan *minimum distance*. Algoritma non-parametrik menggunakan asumsi distribusi bebas, menjadi keunggulannya yang akan lebih baik

digunakan pada distribusi kelas dengan variabilitas yang lebar. Terdapat beberapa algoritma non-parametrik, antara lain adalah Jaringan Syaraf Tiruan (JST), Pohon Keputusan, Logika Fuzzy, dan *Support Vector Machine* (SVM).

Data-data penginderaan jauh dan Sistem Informasi Geografis (SIG) saat ini telah banyak dijumpai. Sehingga diperlukan suatu metode yang dapat memanfaatkan data-data tersebut untuk memperoleh informasi baru. Teknik *data mining* (*Knowledge Discovery from Data/KDD*) banyak digunakan dalam mengolah data-data yang cukup banyak untuk memperoleh informasi baru yang bermanfaat. Salah satu metode dalam KDD untuk mengekstraksi informasi adalah menggunakan metode *machine learning*. CTA merupakan algoritma *machine learning* yang dapat digunakan untuk mengklasifikasi data penginderaan jauh dan data pendukungnya. Meskipun kadang-kadang disebut sebagai *decision tree* (pohon keputusan), namun lebih cenderung sebagai jenis pohon keputusan yang mengarah ke keputusan kategoris (Clark Labs, 2008). Dalam Wu dan Kumar (2009) algoritma C4.5 dan CART termasuk sepuluh algoritma teratas yang paling berpengaruh dan banyak digunakan dalam *data mining*, dengan algoritma C4.5 urutan pertama dan algoritma CART urutan kesepuluh. Hal ini diungkapkan dalam komunitas *data mining* pada konferensi *Institute of Electrical and Electronics Engineers (IEEE) International Conference in Data Mining (ICDM)* di tahun 2006.

Terdapat keberkaitan penelitian ini dengan penelitian-penelitian yang telah dilakukan sebelumnya, seperti penelitian Li *et al.* (2000), dalam penggunaan teknik *data mining* dalam mengekstraksi informasi dari data penginderaan jauh dan SIG untuk klasifikasi penggunaan lahan. Penelitian Zambon *et al.* (2006), dalam penggunaan metode CTA untuk klasifikasi pada data penginderaan jauh. Berkaitan dengan penelitian Samudra (2007), Indrawati (2009), dan Supriyadi (2014), dalam menganalisis kemampuan metode non-parametrik untuk klasifikasi penutup atau penggunaan lahan melalui beberapa parameter dengan menggunakan citra penginderaan jauh.

Perbedaan penelitian ini dengan penelitian-penelitian sebelumnya adalah dalam penelitian ini mengkaji kemampuan metode non-parametrik *Classification Tree Analysis* (CTA) dengan teknik *data mining* untuk klasifikasi penggunaan lahan menggunakan citra Landsat-8 OLI dengan cara melakukan simulasi kombinasi parameter yang akan menghasilkan akurasi terbaik, serta waktu yang diperlukan dalam membentuk pohon keputusan dan melakukan klasifikasi serta total waktu yang

diperlukan, sehingga efektivitas metode CTA dapat diketahui. Membandingkan parameter tiga aturan pemisah (*splitting rules*) dalam CTA, yaitu *Ratio*, *Entropy*, dan *Gini*. Dan empat parameter pemangkas (*pruning*) pohon keputusan, yaitu 0%, 1%, 5%, dan 10%. Dengan masukan data klasifikasi, memanfaatkan data-data yang bersumber dari citra Landsat-8 OLI dan data ketinggian, serta turunannya dari proses transformasi citra dan filter tekstur, dan data kemiringan lereng dari data ketinggian. Penelitian ini juga membandingkan dua level skema klasifikasi penggunaan lahan yang berbeda.

Semakin banyaknya data seiring dengan berjalannya waktu diperlukan metode untuk membantu menganalisis data-data tersebut menjadi informasi baru yang bermanfaat, dalam hal ini informasi penggunaan lahan. Menggunakan teknik *data mining* atau KDD dalam mengekstraksi informasi penggunaan lahan dari data-data yang ada, dengan cara memanfaatkan CTA sebagai metode *machine learning* pada KDD. Pemilihan aturan pemisah dan pemangkas metode CTA, kombinasi parameter-parameter data masukan dari citra Landsat-8 OLI serta data non-spektral, dan penerapannya pada dua skema klasifikasi penggunaan lahan dari modifikasi skema BPN berbasis penutup lahan yang berbeda tingkat kerinciannya, memerlukan pengujian terkait dengan tingkat akurasi hasil klasifikasi penggunaan lahan yang dihasilkan. Pemanfaatan data citra penginderaan jauh dan pendukungnya menggunakan metode non-parametrik CTA dengan teknik *data mining* untuk klasifikasi penggunaan lahan berbasis penutup lahan dengan dua level klasifikasi perlu dilakukan penelitian untuk mengetahui apakah memenuhi kriteria akurasi yang disyaratkan dalam klasifikasi penggunaan lahan.

Penelitian ini bertujuan melakukan simulasi dari beberapa kombinasi parameter data masukan, parameter CTA, dan parameter skema klasifikasi untuk mengetahui tingkat akurasi hasil klasifikasi penggunaan lahan, serta memperoleh pohon keputusan dari hasil KDD. Serta menganalisis akurasi metode non-parametrik CTA dengan teknik *data mining* untuk klasifikasi penggunaan lahan menggunakan citra Landsat-8 OLI serta menerapkan hasil dari KDD pada daerah lain yang memiliki karakteristik mirip.

METODE PENELITIAN

Penelitian ini akan mengkaji metode CTA dalam menangani data yang banyak untuk klasifikasi penggunaan lahan dan perbandingannya dalam menangani dua tingkat skema klasifikasi yang berbeda.

Dengan melakukan simulasi beberapa parameter metode *splitting rule* dalam CTA, seperti: *information gain*, *gain ratio*, dan *Gini index*; dan parameter pemangkasan pohon (*tree pruning*) yang berbeda, yakni 0%, 1%, 5%, dan 10%; serta penerapannya pada dua tingkat skema klasifikasi yang berbeda. Kemudian mencatat waktu proses dan akurasi. Setelah mendapatkan kombinasi parameter dengan akurasi terbaik pada setiap skema klasifikasi penggunaan lahan, kemudian hasil klasifikasi tersebut disusun menjadi peta penggunaan lahan. Pohon keputusan yang diperoleh dari kombinasi parameter terbaik tersebut lalu diterapkan pada daerah lain yang memiliki karakteristik mirip.

Terdapat 34 data masukan yang akan digunakan dalam penelitian ini, yang diperoleh dari: citra Landsat-8 OLI tujuh saluran; filter tekstur *variance* dan tekstur *mean*; transformasi *Normalized Difference Vegetation Index* (NDVI), *Normalized Difference Water Index* (NDWI), *Bare soil Index* (BI), *Normalized Difference Built-up Index* (NDBI), dan *Principal Component Analysis* (PCA); data ketinggian dan kemiringan lereng. Data ketinggian dan kemiringan lereng, diperoleh dari peta Rupa Bumi Indonesia (RBI) skala 1:25.000 tahun 2000.

Penentuan sampel dalam penelitian ini mempertimbangkan kelas-kelas yang ada dalam skema klasifikasi dan distribusinya dalam lokasi penelitian. Kemudian sampel tersebut akan dilakukan verifikasi dengan melakukan cek lapangan, jika ada sampel yang kurang tepat dapat dibetulkan kembali. Dalam melakukan sampling penelitian ini akan menggunakan metode *purposive random sampling*. Dimana metode *purposive random sampling*, mengambil sampel dari beberapa unit yang kecil/sedikit secara *random* (acak) dari keseluruhan target yang lebih luas untuk tujuan spesifik/tertentu dalam penelitian (Teddlie & Yu, 2007). Menurut Saleh (2014) teknik pengambilan sampel dengan metode *purposive random sampling*, yaitu sampel diambil secara acak berdasarkan tujuan yang telah ditetapkan dengan asumsi dan pertimbangan sampel telah mewakili kelas penggunaan lahan yang telah ditentukan serta kemudahan akses untuk mencapai sampel.

Dari sampel tersebut kemudian dipisahkan antara sampel yang akan digunakan untuk data *training* dan data uji akurasi. Uji akurasi yang dalam penelitian ini menggunakan *overall accuracy* (akurasi keseluruhan), *producers accuracy* (akurasi pembuat), *users accuracy* (akurasi pengguna), dan *kappa* (K).

HASIL DAN PEMBAHASAN

Data

Bahan-bahan yang diperlukan dan data yang digunakan dalam penelitian ini antara lain adalah sebagai berikut: 1) Citra satelit Landsat-8 OLI yang direkam pada tanggal 24 Juni 2013 (LC81200652013175LGN00), *path* 120 dan *row* 65, yang telah terkoreksi Level 1-T (*Standard Terrain Correction*); 2) Citra Digital Globe Google Earth daerah Kabupaten Purworejo dan Kabupaten Kebumen dengan tanggal perekaman adalah 27 Februari 2014, 31 Mei 2014, 3 Juni 2014, 11 Juni 2014, 18 Agustus 2014 dan 19 Agustus 2014; 3) Peta Rupa Bumi Indonesia (RBI) skala 1:25.000 Tahun 2000; 4) Skema klasifikasi penggunaan lahan yang merujuk pada klasifikasi penggunaan tanah Norma Standar Prosedur Kriteria Pemetaan Tematik Badan Pertanahan Nasional (NSPK-BPN) tahun 2012 yang dimodifikasi sesuai dengan informasi yang dapat diinterpretasi melalui citra satelit dan disesuaikan dengan kondisi daerah penelitian.

Data Cleaning

Tahapan awal penelitian ini adalah koreksi radiometrik citra Landsat-8OLI. Koreksi radiometrik diperlukan untuk memperbaiki nilai piksel supaya sesuai dengan yang seharusnya. Dalam penelitian ini akan menggunakan koreksi radiometrik kalibrasi citra *reflectance* Landsat-8 menggunakan data yang terdapat dalam *header* citra pada file metadata (MTL). Hal ini diperlukan karena akan menggunakan data hasil dari transformasi citra, sehingga diharapkan memperoleh nilai transformasi yang tepat. Lokasi penelitian dapat dilihat pada Gambar 1. Pemilihan lokasi tersebut dikarenakan pada daerah ini memiliki variasi penggunaan lahan yang cukup baik dan terdapat daerah yang datar maupun bergunung, serta diharapkan dalam melakukan proses klasifikasi tidak memerlukan waktu yang cukup lama karena memiliki luas yang lebih kecil.

Koreksi geometrik pada penelitian ini menggunakan acuan peta Rupa Bumi Indonesia (RBI) skala 1:25.000, dengan menggunakan koreksi *image to map registration*, dengan sistem proyeksi *Universal Transverse Mercator* (UTM), datum WGS 1984, dan zona 49 South. Penelitian ini menggunakan koreksi geometrik orde 3, dengan mempertimbangkan variasi topografi dimana terdapat daerah yang datar, landai, dan bergunung. Jumlah GCP dalam penelitian ini sebanyak 25 titik, menghasilkan nilai *Root Mean Square Error* (RMSE) sebesar 0,27. Untuk tetap menjaga nilai piksel agar sesuai dengan yang sebenarnya, maka menggunakan metode interpolasi algoritma *nearest neighbour*. Algoritma *nearest neighbour* diterapkan

dengan mengambil kembali nilai dari piksel terdekat yang telah tergeser ke posisi baru.

Data Integration dan Data Transformation

Penelitian ini akan menggunakan 7 (tujuh) saluran multispektral citra Landsat-8 OLI, dengan tidak menyertakan saluran *cirrus* (*band* 9) dan pankromatik (*band* 8). Dalam membantu agar klasifikasi memperoleh hasil klasifikasi yang akurat, maka memanfaatkan transformasi citra yang dapat menonjolkan suatu indeks. Indeks tersebut diharapkan dapat berkorelasi dengan kelas tertentu. Transformasi *Normalized Difference Vegetation Index* (NDVI) digunakan untuk menonjolkan vegetasi, dari transformasi tersebut menghasilkan nilai -0,391686 sampai dengan 0,832692 yang melibatkan saluran inframerah dekat (*band* 5) dan saluran merah (*band* 4). *Normalized Difference Water Index* (NDWI) yang menonjolkan air diharapkan dapat membantu dalam menentukan kelas yang terdapat unsur air. Penerapan transformasi NDWI pada citra Landsat-8 OLI penelitian ini menghasilkan nilai -0,594049 sampai dengan 0,884235 yang menggunakan saluran hijau (*band* 3) dan saluran inframerah tengah 1 (*band* 6). Dengan menggunakan *Bare soil Index* (BI) yang menonjolkan tanah terbuka, menghasilkan nilai -3,171766 sampai dengan 1,348825 yang melibatkan empat saluran, yaitu: saluran inframerah tengah 1 (*band* 6), saluran merah (*band* 4), saluran inframerah dekat (*band* 5), dan saluran biru (*band* 2). Hasil dari transformasi *Normalized Difference Built-up Index* (NDBI), memiliki nilai antara -0,827258 sampai dengan 0,318246. NDBI menggunakan saluran inframerah tengah 1 (*band* 6) dan saluran inframerah dekat (*band* 5).

Penelitian ini juga akan menggunakan transformasi *Principal Component Analysis* (PCA). Hasil dari ketujuh saluran citra Landsat-8 menghasilkan tujuh PC dari proses transformasi PCA. Ketujuh PC hasil transformasi PCA akan dimasukkan sebagai data masukan penelitian. Meskipun PC7 nampak banyak *noise*, namun tetap akan dimasukkan dalam proses *training*. CTA akan menentukan apakah saluran tersebut akan digunakan sebagai *node* (simpul keputusan) atau tidak digunakan dalam proses klasifikasi.

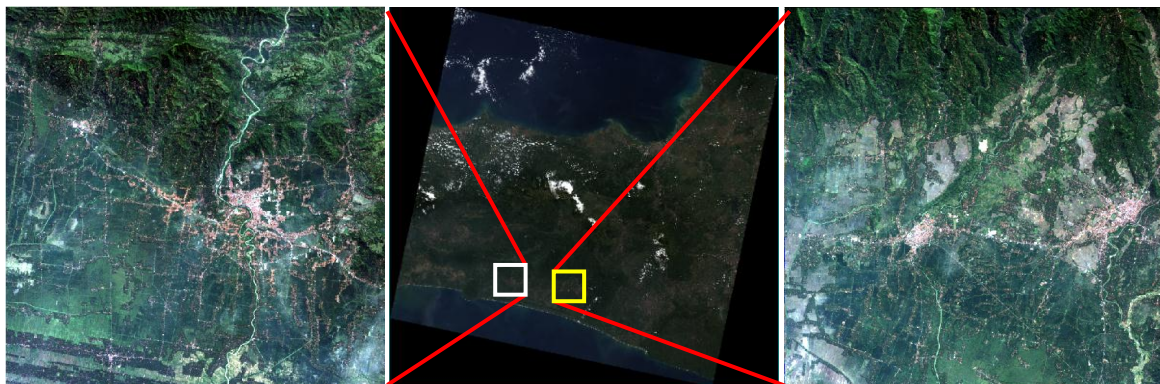
Pemanfaatan filter tekstur dalam penelitian ini diharapkan batas antarkelas maupun kesamaan kelas akan terbantu dalam proses klasifikasi menggunakan CTA. Filter tekstur *Mean* digunakan untuk memperhalus kenampakan citra. Penghalusan gradasi nilai dari satu piksel ke piksel lain sehingga perubahan nilai yang mencolok dapat ditekan dengan memberikan nilai baru yang merupakan rerata dari

nilai-nilai piksel tetangga. Dalam penelitian ini akan menggunakan filter tekstur *mean occurance* dengan *processing window* 3x3 dan 5x5. Penggunaan tekstur *mean* diharapkan dapat membantu klasifikasi dalam menentukan kelas yang sama dari nilai tekstur *mean* disekitarnya atau dari nilai rerata nilai piksel tetangga. Filter tekstur *variancemengukur* homogenitas lokal dari suatu citra. Tekstur *variance occurance* dengan *processing window* 3x3 dan 5x5 digunakan dalam penelitian ini sebagai data masukan. Filter tekstur *variance* ini diharapkan dapat membantu dalam memisahkan antar kelas yang berbeda dari nilai variasi nilai piksel tetangga. Data non-spektral yang digunakan dalam penelitian ini adalah data ketinggian dan data kemiringan lereng. Data ketinggian diperoleh dari data kontur peta RBI 1:25.000.

Untuk melakukan klasifikasi dari citra menjadi kelas-kelas penggunaan lahan, suatu algoritma klasifikasi memerlukan pelatihan untuk mengenali kelas-kelas tersebut. Tabel 1 merupakan skema klasifikasi penggunaan lahan penelitian yang mengacu

dari skema klasifikasi penggunaan tanah BPN-RI tahun 2012 berbasis penutup lahan yang dapat diinterpretasi dari citra Landsat-8 OLI pada daerah penelitian di sebagian wilayah Kabupaten Purworejo. Melakukan interpretasi citra untuk mengidentifikasi objek yang digunakan sebagai daerah sampel. Dengan melakukan komposit citra, baik menggunakan *true colour* (komposit saluran 432) maupun *false colour*, seperti komposit saluran 563 dan saluran 543. Selain itu dalam mengidentifikasi objek juga menggunakan bantuan dari Citra Digital Globe Google Earth yang memiliki resolusi lebih tinggi.

Hasil dari daerah sampel ini kemudian dilakukan cek lapangan untuk memastikan daerah sampel tersebut telah sesuai dengan keadaan (kelas) yang sebenarnya di lapangan. Jika terdapat sampel yang tidak sesuai dapat dibetulkan kembali. Daerah sampel ini juga dipisahkan/dibedakan, daerah sampel yang digunakan untuk *training area* (daerah contoh) dan daerah sampel yang digunakan untuk *accuracy assesment* (uji akurasi).



Gambar 1. Citra Lokasi Penelitian Kabupaten Purworejo (kotak kanan) dan Kabupaten Kebumen (kotak kiri)

Tabel 1. Skema Klasifikasi Penggunaan Lahan Penelitian

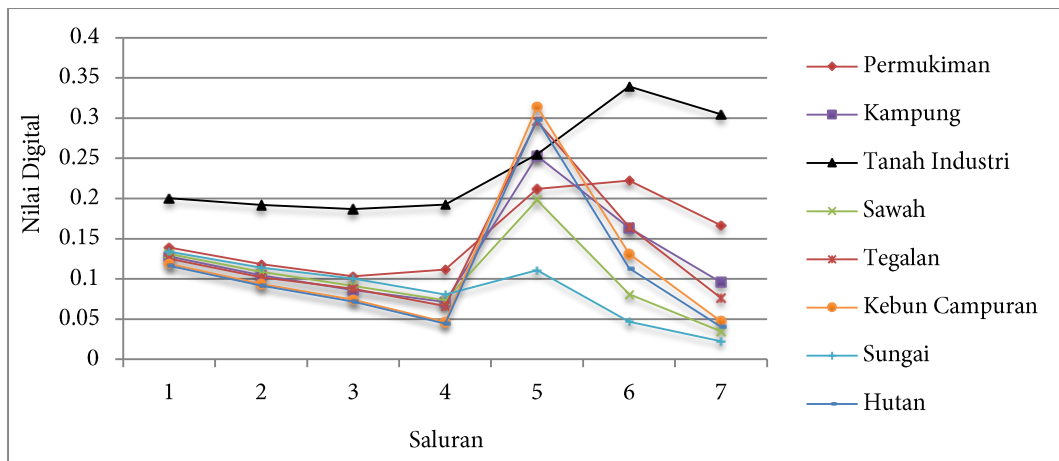
No	Klasifikasi P.L Level I	Kode	No	Klasifikasi P.L Level II	Kode
1	§ Perkampungan	Pk	1	§ Perumahan	Pr
2	§ Tanah Industri/ Pergudangan	Ti	2	§ Kampung	Kp
3	§ Tanah Pertanian	Tp	3	§ Tanah Industri/ Pergudangan	Ti
4	§ Perairan Darat	Pd	4	§ Sawah	Sw
5	§ Hutan	Hn	5	§ Tegalan/Ladang	Tg
			6	§ Kebun Campuran	Kc
			7	§ Sungai	Su
			8	§ Hutan	Hn

Data Selection

Panduan paling penting dalam memilih lokasi sampel untuk data *training* adalah memastikan bahwa semua variabilitas dalam kelas tercatat. Untuk kelas yang homogen hanya memerlukan beberapa lokasi sampel, sedangkan kelas dengan variabilitas tinggi

memerlukan lokasi sampel yang lebih banyak (McCoy, 2005).

Penelitian ini menggunakan 86 sampel daerah contoh dan 61 sampel uji akurasi untuk kabupaten purworejo, serta 64 sampel uji akurasi untuk kabupaten kebumen. Gambar 2 merupakan sebaran sampel daerah contoh dan sebaran sampel uji akurasi.



Gambar 3. Grafik Mean Sampel Training Area Kelas Penggunaan Lahan

Data Mining

Penelitian ini menggunakan perangkat lunak IDRISI Selva 17.02 untuk proses klasifikasi. Data masukan yang digunakan dalam penelitian ini sebanyak 34, dan satu untuk data *training*. Data *training area* (daerah contoh) dalam penelitian ini terdiri atas dua data, yaitu untuk data Level I yang terdiri atas 5 kelas dan untuk data Level II yang terdiri atas 8 kelas yang lebih rinci. Parameter-parameter klasifikasi terlebih dahulu dimasukkan dan dipilih sebelum melakukan eksekusi klasifikasi.

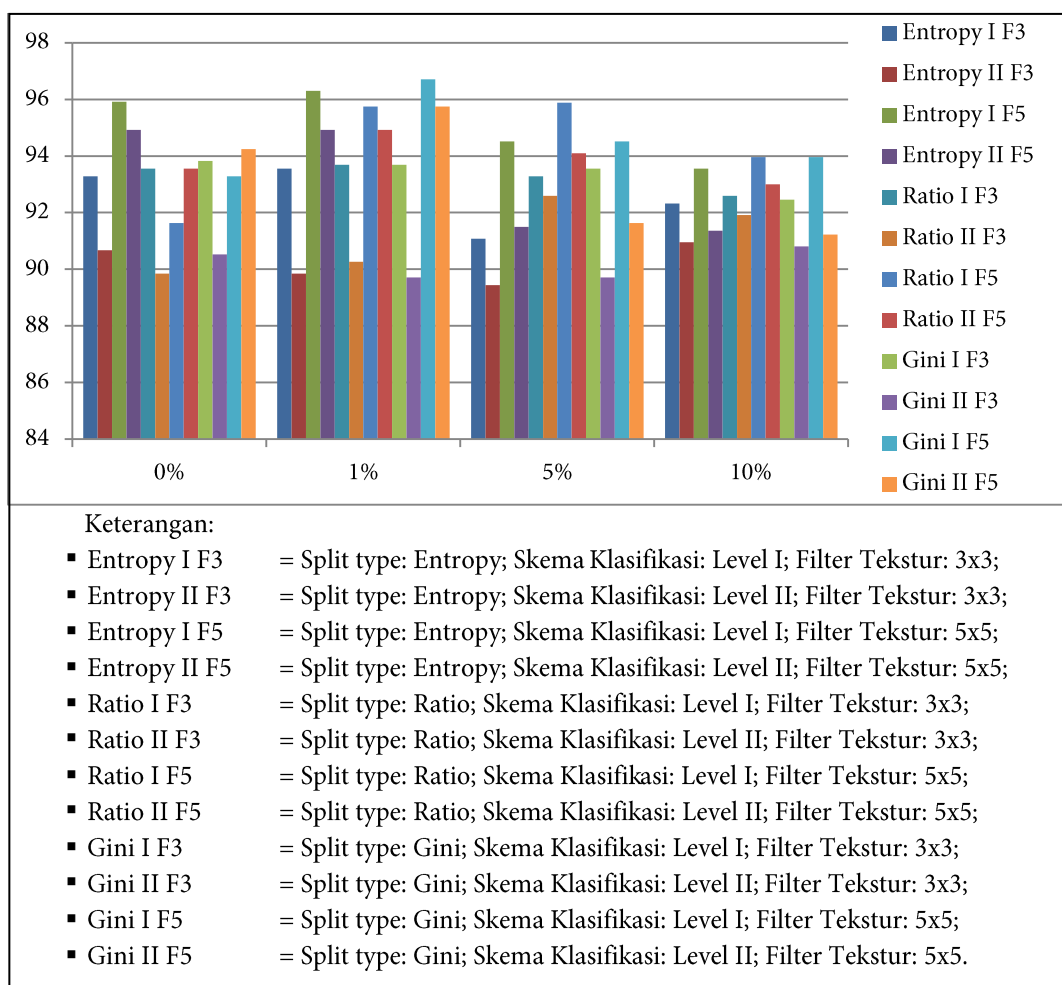
Simulasi klasifikasi penggunaan lahan menggunakan metode *Classification Tree Analysis* (CTA) yang dilakukan dalam penelitian ini sebanyak 48 simulasi, dengan beberapa kombinasi parameter. Dari hasil simulasi tersebut kemudian dilakukan uji akurasi menggunakan data *accuracy assessment* (uji akurasi).

Pattern Evaluation

Pencatatan waktu dalam penelitian ini menggunakan stopwatch, sehingga akan terdapat selisih waktu beberapa mili-detik, berkaitan dengan respon peneliti ketika mengamati proses dengan penghentian stopwatch. Proses algoritma CTA pada umumnya tergolong cepat dengan hanya memerlukan waktu terlama eksekusi 29,81 detik untuk mengolah 34 data masukan dengan 8 kelas penggunaan lahan serta tanpa ada pemotongan pohon keputusan, dan menggunakan algoritma *Ratio*.

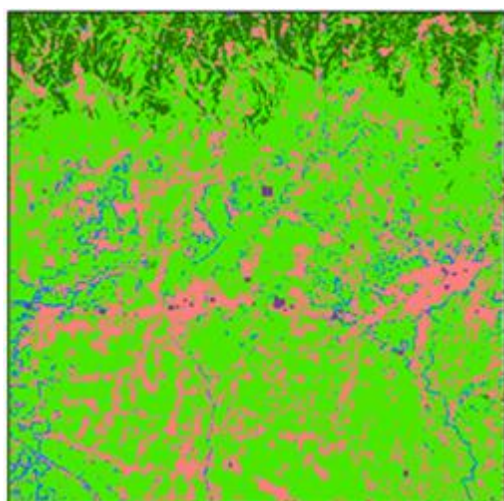
Rata-rata waktu eksekusi simulasi adalah selama 7 detik. Hal yang perlu diperhatikan adalah perangkat keras dalam pemrosesan juga mempengaruhi waktu proses. Waktu proses yang cepat ini memang termasuk keunggulan dari metode CTA, yaitu efisiensi komputasi. Gambar 4 merupakan grafik akurasi hasil simulasi. Pada Gambar 5 dapat dilihat bagaimana hasil klasifikasi terakurat pada setiap parameter pemisah. Dari ketiga atribut pemisah CTA, *Ratiomemiliki* rata-rata akurasi tertinggi, yaitu 93,16% dan *kappa* 0,9094, urutan kedua adalah *Gini* dengan rata-rata 92,85% dan *kappa* 0,9063, dan ketiga adalah *Entropy* dengan rata-rata 92,76% dan *kappa* 0,9042.

Akurasi keseluruhan maksimum parameter *Ratio* adalah 95,88% dan akurasi minimum 89,85%. Pada parameter *Gini*, akurasi maksimum sebesar 96,71% dan akurasi minimum sebesar 89,71%. Untuk parameter *Entropy*, akurasi maksimum sebesar 96,3% dan akurasi minimum 89,44%. Parameter *pruning* yang memiliki rata-rata akurasi terbaik adalah 1% dengan nilai 93,76% dan *kappa* 0,9185, kedua dengan nilai rata-rata 92,94% dan *kappa* 0,9056 adalah 0%, urutan ketiga dengan nilai rata-rata 92,65% dan *kappa* 0,9036 adalah 5%, dan yang terakhir 10% dengan rata-rata akurasi 92,34% dan *kappa* 0,8988. Sedangkan pada filter tekstur dengan jendela bergerak 5x5 memiliki rata-rata akurasi lebih tinggi dibandingkan dengan jendela bergerak 3x3, rata-rata akurasi jendela bergerak 5x5 adalah 94,04% dan *kappa* 0,9208 dan rata-rata akurasi jendela bergerak 3x3 adalah 91,8% dan *kappa* 0,8924.



Gambar 4. Grafik Akurasi Hasil Klasifikasi

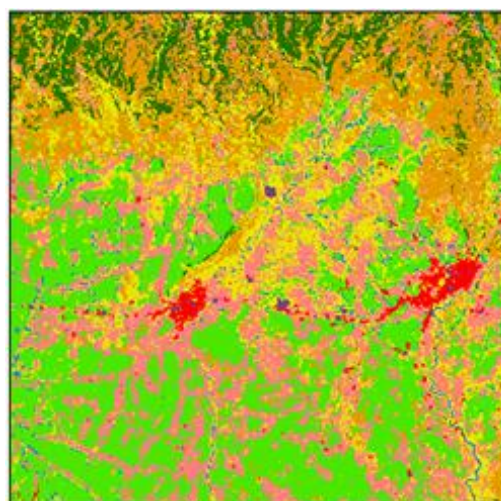
Skema Klasifikasi Penggunaan Lahan Level I



- Keterangan:**
- Perkampungan (Pk)
 - Tanah Industri (Ti)
 - Tanah Pertanian (Tp)
 - Perairan Darat (Pd)
 - Hutan (Hn)

Skema Klasifikasi Penggunaan Lahan Level II

Entropy

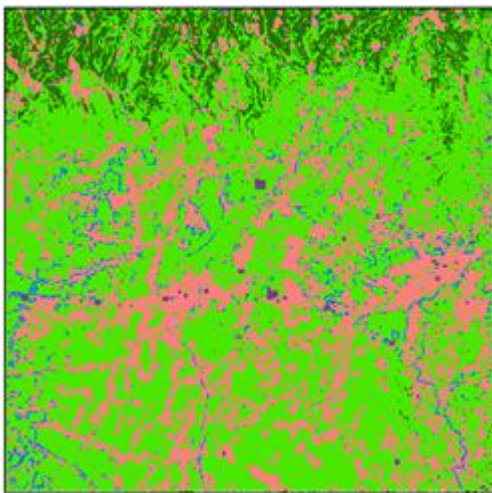


- Keterangan:**
- Perumahan (Pr)
 - Kampung (Kp)
 - Tanah Industri (Ti)
 - Sawah (Sw)
 - Tegalan (Tg)
 - Kebun Campuran (Kc)
 - Sungai (Su)
 - Hutan (Hn)

Gambar 5. Hasil Klasifikasi Terakurat pada Parameter Pemisah

Lanjutan Gambar 5.

Skema Klasifikasi Penggunaan Lahan Level I

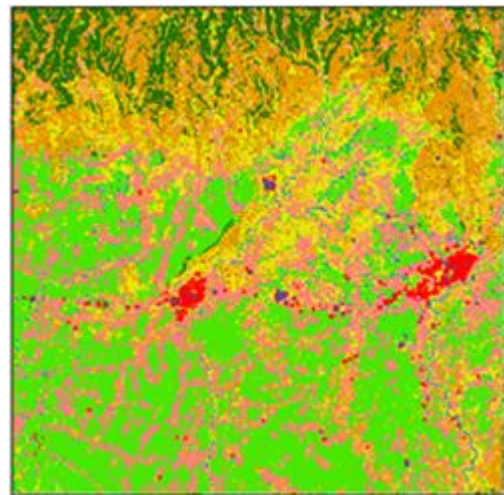


Keterangan:

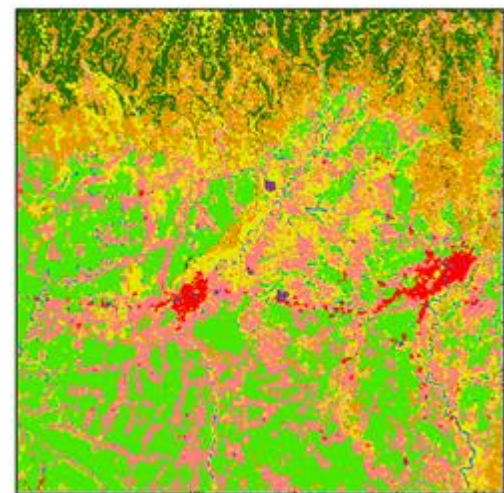
Perkampungan (Pk)	Perairan Darat (Pd)
Tanah Industri (Ti)	Hutan (Hn)
Tanah Pertanian (Tp)	

Skema Klasifikasi Penggunaan Lahan Level II

Ratio



Gini



Keterangan:

Perumahan (Pr)	Tegalan (Tg)
Kampung (Kp)	Kebun Campuran (Kc)
Tanah Industri (Ti)	Sungai (Su)
Sawah (Sw)	Hutan (Hn)

Skema klasifikasi penggunaan lahan yang lebih general dengan 5 kelas (Level I) memiliki rata-rata akurasi lebih tinggi dibandingkan dengan skema klasifikasi penggunaan lahan yang lebih rinci dengan 8 kelas (Level II). Pada skema Level I rata-rata akurasi adalah 93,87% dan $kappa$ 0,9072, sedangkan pada skema Level II adalah 91,97% dan $kappa$ 0,9060.

Hasil klasifikasi penggunaan lahan dengan akurasi terbaik dalam penelitian ini diperoleh simulasi CTA18 dengan 5 kelas penggunaan lahan (skema Level I). Gambar 6 menunjukkan bagaimana pohon keputusan mengambil keputusan mengalir dari akar melewati cabang sampai ke daun. Data masukan yang berperan dalam pembentukan klasifikasi penggunaan lahan Level I terbaik ini antara lain adalah:

1. Band 34 = Filter Tektur Varian 5x5 Band 7;
2. Band 12 = Kemiringan Lereng;
3. Band 16 = PCA PC3;
4. Band 5 = Citra Landsat-8 Saluran 5;
5. Band 4 = Citra Landsat-8 Saluran 4;
6. Band 1 = Citra Landsat-8 Saluran 1;
7. Band 8 = Bare-soil Index (BI);
8. Band 18 = PCA PC5;
9. Band 21 = Filter Tektur Mean 5x5 Band 1;
10. Band 28 = Filter Tektur Varian 5x5 Band 1;
11. Band 11 = *Normalized Difference Water Index* (NDWI);
12. Band 17 = PCA PC4;
13. Band 23 = Filter Tektur Mean 5x5 Band 3;
14. Band 14 = PCA PC1;

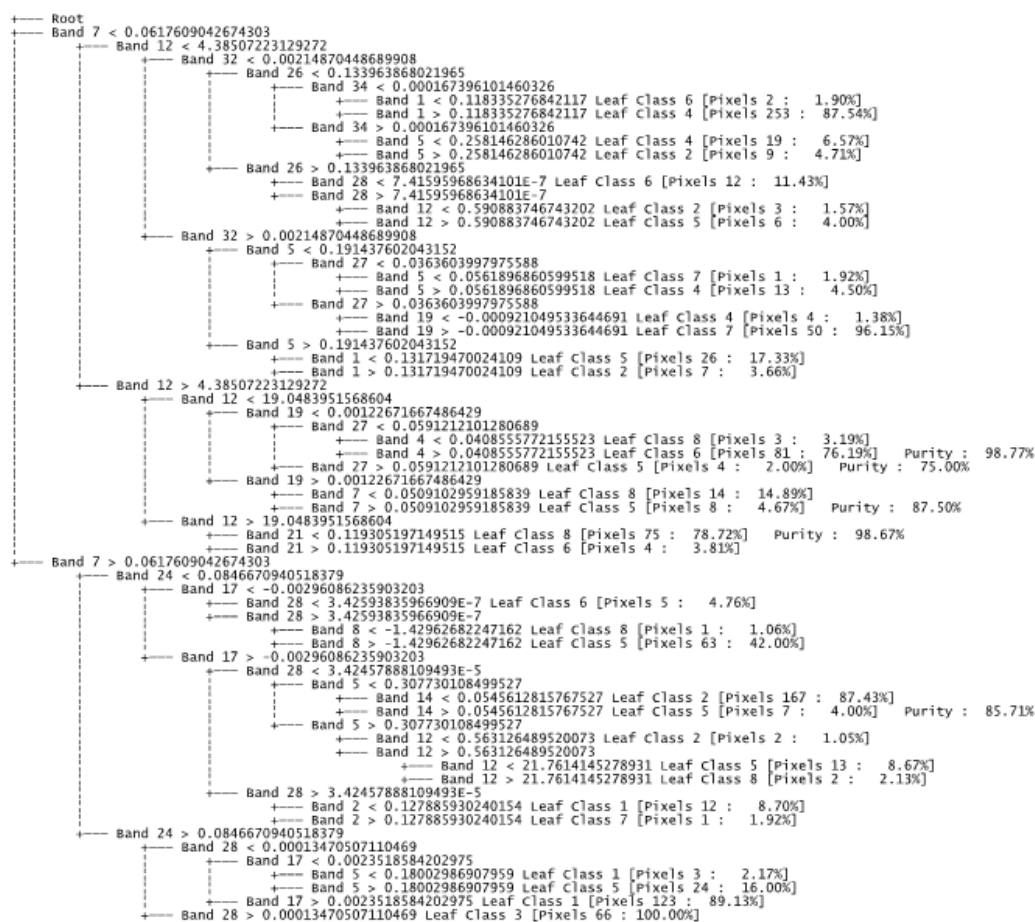


Gambar 6. Pohon Keputusan Klasifikasi Skema Level I Terbaik

Kompleksitas pohon biasanya diukur dengan salah satu dari beberapa hal berikut: jumlah *node*, jumlah *leaf*, *tree depth* (kedalaman pohon) dan jumlah atribut yang digunakan. Jumlah *node* pada pohon keputusan klasifikasi terbaik skema level I adalah 22. Jumlah *leaf* adalah 22, kedalaman pohon keputusan adalah 8. Dari pohon keputusan tersebut paling sederhana adalah dalam menentukan kelas Tanah Industri (Ti) atau dalam Gambar 6 berupa "Class 2" dimana hanya memerlukan 2 *node*, yaitu Band 34 dan Band 28. Sedangkan yang memerlukan banyak *node* adalah dalam memisahkan kelas Tanah Pertanian (Class 3) dengan Hutan (Class 5). Pada skema klasifikasi penggunaan lahan Level II, akurasi terbaik diperoleh simulasi CTA42. Alur proses penentuan 8 kelas penggunaan lahan (skema Level II) oleh pohon keputusan dapat dilihat pada Gambar 7. Data masukan

yang berperan dalam menentukan dan memisahkan masing-masing kelas adalah sebagai berikut :

1. Band 7 = Citra Landsat-8 Saluran 7;
2. Band 12 = Kemiringan Lereng;
3. Band 32 = Filter Tektur Varian 5x5 Band 5;
4. Band 26 = Filter Tektur Mean 5x5 Band 6;
5. Band 34 = Filter Tektur Varian 5x5 Band 7;
6. Band 1 = Citra Landsat-8 Saluran 1;
7. Band 5 = Citra Landsat-8 Saluran 5;
8. Band 28 = Filter Tektur Varian 5x5 Band 1;
9. Band 27 = Filter Tektur Mean 5x5 Band 7;
10. Band 19 = PCA PC6;
11. Band 4 = Citra Landsat-8 Saluran 4;
12. Band 21 = Filter Tektur Mean 5x5 Band 1;
13. Band 24 = Filter Tektur Mean 5x5 Band 4;
14. Band 17 = PCA PC4;
15. Band 8 = *Bare-soil Index* (BI);
16. Band 14 = PCA PC1;
17. Band 2 = Citra Landsat-8 Saluran 2;



Gambar 7. Pohon Keputusan Klasifikasi Skema Level II Terbaik

Jumlah *node* pada pohon keputusan klasifikasi terbaik skema level II adalah 33. Jumlah *leaf* adalah 34, kedalaman pohon keputusan adalah 7. Dari pohon keputusan klasifikasi terbaik skema level II yang paling sederhana adalah dalam menentukan kelas Tanah Industri (Ti) atau berupa “Class 3” dimana hanya memerlukan 3 *node*, yaitu Band 7, Band 24 dan Band 28. Sedangkan yang memerlukan banyak *node* adalah dalam memisahkan kelas Tegalan (Class 5) dengan kelas yang lainnya.

Dari kedua pohon keputusan di atas dapat diketahui bahwa untuk menentukan kelas penggunaan lahan skema Level I memerlukan 14 saluran data masukan dari 34 data masukan yang diberikan. Sedangkan untuk menentukan kelas penggunaan lahan skema Level II memerlukan 17 saluran data masukan atau separuh dari 34 data masukan yang diberikan dalam penelitian.

Knowledge Presentation

Hasil dari proses KDD berupa pengetahuan, dalam hal ini informasi penggunaan lahan. Dari proses *pattern evaluation*, sehinggadiketahui pohon keputusan yang memiliki akurasi terbaik. Penelitian ini menghasilkan pohon keputusan terbaik dari CTA18 (Level I) dan CTA42 (Level II), hasil klasifikasi dari

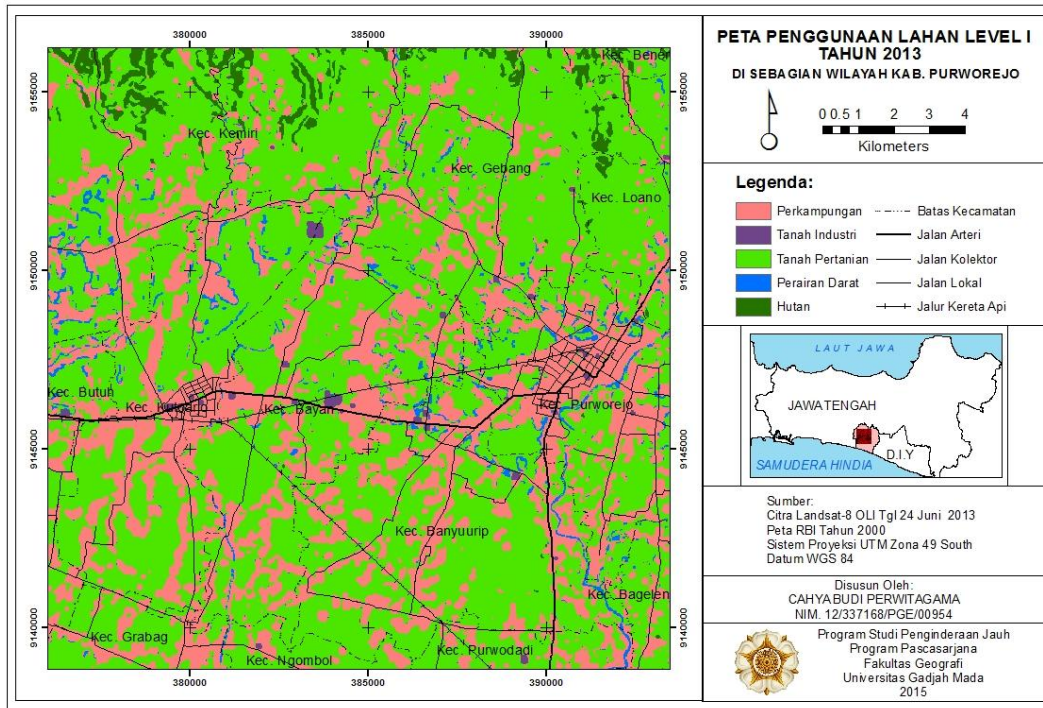
pohon keputusan tersebut kemudian dilanjutkan ke proses selanjutnya dalam KDD yaitu *knowledge presentation*. Untuk merepresentasikan atau memvisualisasikan pengetahuan dari pengetahuan yang telah ditambang, maka dalam penelitian ini pengetahuan tersebut disajikan dalam bentuk peta, yaitu Peta Penggunaan Lahan Level I dan Level II tersebut dapat dilihat pada Gambar 8.

Penerapan Hasil KDD pada Daerah Lain

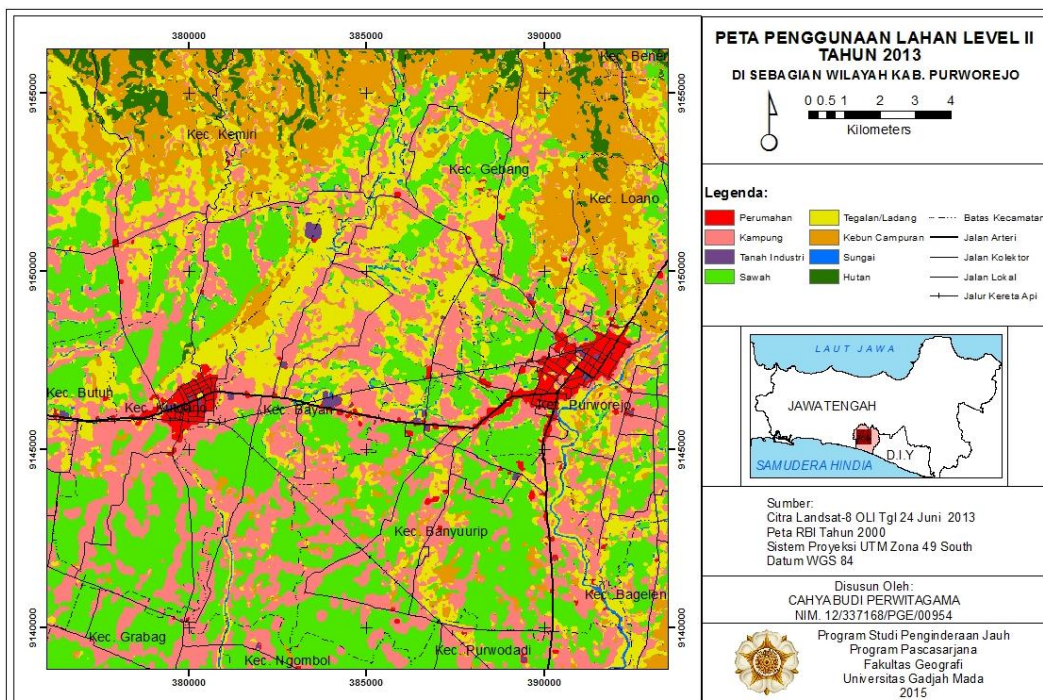
Hasil pohon keputusan dari *training* sebelumnya, dapat dimanfaatkan pada daerah lain yang memiliki kemiripan karakteristik. Dalam penelitian ini akan mengambil daerah di sebelah barat Kabupaten Purworejo yang memiliki kemiripan karakteristik, yaitu Kabupaten Kebumen. Pada data Kabupaten Kebumen sebelumnya juga dilakukan data *pre-processing* terlebih dahulu, kemudian data-data baru tersebut dapat digunakan sebagai masukan dari pohon keputusan. Untuk mengetahui bagaimana tingkat akurasi klasifikasi penggunaan lahan pada daerah lain menggunakan pohon keputusan hasil pembelajaran, maka diperlukan juga pengambilan sampel uji akurasi. Jumlah sampel yang digunakan untuk uji akurasi di sebagian wilayah Kabupaten Kebumen sebanyak 64 sampel.

Dalam mengeksekusi pohon keputusan dari *learning* (pembelajaran) sebelumnya, penelitian ini menggunakan software ENVI 5.1 menggunakan modul *Decision Tree*. Penyusunan pohon keputusan sesuai dengan yang diperoleh dari pembelajaran pada IDRISI Selva 17.02. Pohon keputusan yang telah dibentuk tersebut kemudian diberi masukan parameter-

parameter yang digunakan pada setiap *node*, sesuai dengan daerah baru yang akan dilakukan klasifikasi. Pohon keputusan tersebut kemudian dieksekusi, sehingga menjadi klasifikasi penggunaan lahan. Hasil dari klasifikasi penggunaan lahan pada daerah lain, dalam hal ini di sebagian wilayah Kabupaten Kebumen dapat dilihat pada Gambar 9.



(a)

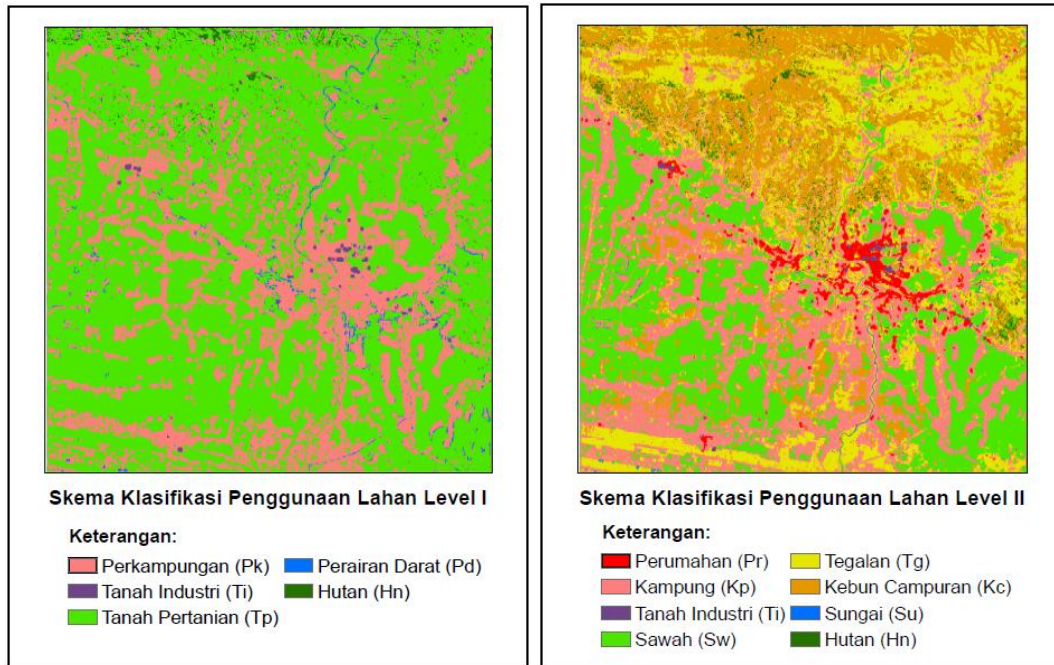


(b)

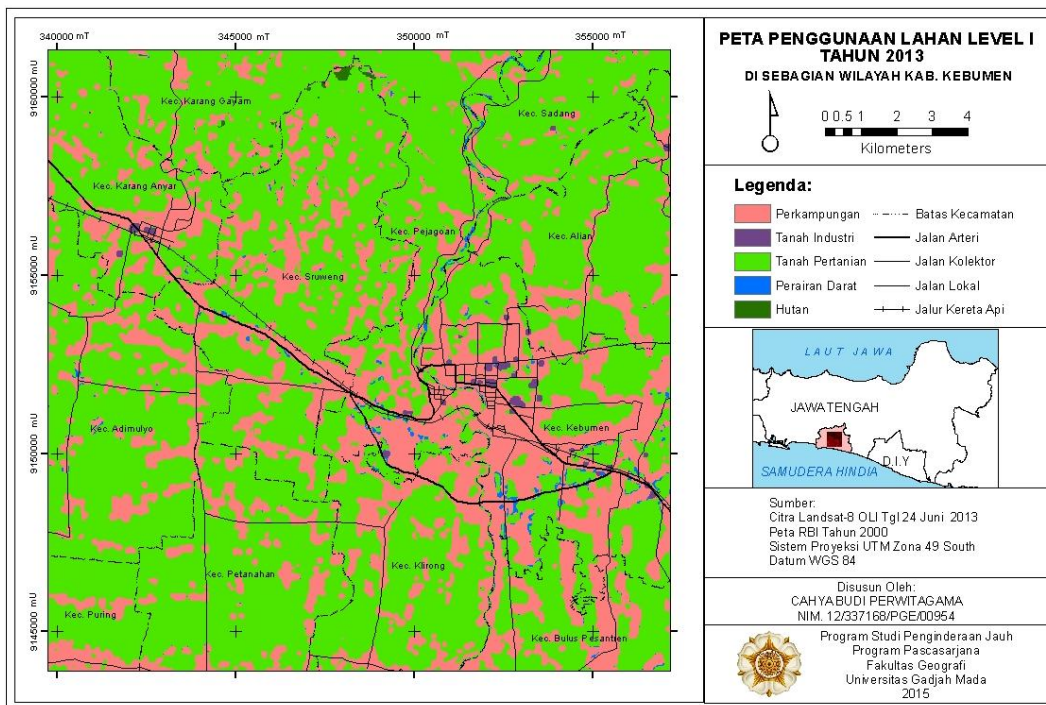
Gambar 8. (a) Peta Penggunaan Lahan Level I dan (b) II Tahun 2013 Kabupaten Purworejo

Hasil dari klasifikasi penggunaan lahan dengan skema penggunaan lahan level I pada daerah lain ini kemudian perlu diketahui tingkat akurasi menggunakan matriks kesalahan. Akurasi keseluruhan (*overall accuracy*) pada hasil klasifikasi dengan skema penggunaan lahan level I diperoleh sebesar 93,27% dan *kappa* 0,8923. Hasil klasifikasi dengan skema penggunaan lahan 8 kelas (level II) dapat diketahui

bahwa besar akurasi keseluruhan (*overall accuracy*) pada hasil klasifikasi dengan skema penggunaan lahan level II diperoleh sebesar 91,36% dan *kappa* 0,8979. Penyajian hasil KDD pada daerah lain, yaitu di sebagian wilayah Kabupaten Kebumen, berupa Peta Penggunaan Lahan Tahun 2013 yang dapat dilihat pada Gambar 10.



Gambar 9. Hasil Klasifikasi Penggunaan Lahan Level I dan Level II Kabupaten Kebumen



Gambar 10. Peta Penggunaan Lahan Level I dan II Tahun 2013 Kabupaten Kebumen

KESIMPULAN DAN SARAN

Kesimpulan yang dapat diambil dari penelitian ini antara lain adalah sebagai berikut:

1. Pengaruh parameter-parameter terhadap hasil klasifikasi dalam penelitian, antara lain adalah sebagai berikut:
 - a. Kombinasi parameter terbaik terdapat pada skema klasifikasi Level I (5 kelas) adalah atribut pemisah *Gini*, *pruning* 1%, dan filter tekstur dengan jendela bergerak 5x5, yaitu dengan akurasi 96,71%, κ 0,9504. Untuk skema klasifikasi penggunaan lahan Level II (8 kelas) kombinasi parameter terbaik adalah atribut pemisah *Gini*, *pruning* 1%, dan filter tekstur 5x5, dengan nilai akurasi 95,75%, κ 0,95.
 - b. *Ratio* menjadi parameter pemisah paling stabil dengan rata-rata akurasi 93,16%, maksimum 95,88%, dan minimum 89,85%. Parameter pemisah yang memiliki akurasi terbaik (tertinggi) adalah *Gini*. Rata-rata proses klasifikasi yang memerlukan waktu lebih lama adalah pada parameter pemisah *Ratio*.
 - c. Skema klasifikasi dengan kelas yang lebih detail dalam hal ini 8 kelas penggunaan lahan (skema P.L Level II) memiliki akurasi lebih rendah jika dibandingkan dengan skema klasifikasi dengan 5 kelas penggunaan lahan (skema P.L Level I).
 - d. Parameter *pruning* dengan rata-rata akurasi lebih tinggi adalah 1%.
Filter tekstur dengan jendela bergerak 5x5 memiliki rata-rata tingkat akurasi lebih tinggi jika dibandingkan dengan jendela bergerak 3x3. Tidak semua data masukan yang diberikan dimanfaatkan oleh CTA dalam proses klasifikasi, dari 34 data masukan yang diberikan untuk Skema P.L Level I dengan akurasi terbaik menggunakan 14 data masukan sedangkan Skema P.L Level II menggunakan 17 data masukan sebagai simpul pemisah.
2. Rata-rata akurasi keseluruhan simulasi CTA sebesar 92,92% dan κ 0,9066, akurasi maksimum metode CTA sebesar 96,71%, κ 0,9504, dan akurasi minimum 89,44%, κ 0,867. Rata-rata waktu proses eksekusi sebesar 6,913 detik, sehingga metode CTA memerlukan waktu proses yang cukup cepat untuk mengolah 34 data masukan dengan ukuran 700x700 piksel. Hasil dari KDD yang diterapkan pada daerah lain dengan karakteristik mirip, menghasilkan akurasi keseluruhan 93,27% dengan κ 0,8923 untuk Skema P.L Level I, sedangkan untuk Skema P.L Level II menghasilkan akurasi keseluruhan 91,36%

dengan κ 0,8979. Dari tingkat akurasi terbaik yang diperoleh dalam penelitian ini maupun penerapan hasil pembelajarannya pada daerah lain, dimana lebih besar dari 90% yang menjadi persyaratan minimal untuk terapan kebijakan penggunaan lahan, maka metode non-parametrik CTA dengan teknik *data mining* menggunakan citra Landsat-8 OLI dapat digunakan sebagai alternatif metode untuk klasifikasi penggunaan lahan berbasis penutup lahan setara dengan skala 1:100.000.

3. Hipotesis pertama dalam penelitian ini tidak terbukti, parameter pemisah *ratio* tidak menghasilkan akurasi terbaik dalam penelitian ini. Karena dalam penelitian ini parameter pemisah *gini* yang memperoleh akurasi terbaik. Hal ini dimungkinkan karena pada parameter pemisah *ratio* sedikit mengalami *overfitting* dibandingkan dengan parameter *gini* pada akurasi terbaiknya. Walaupun demikian parameter *ratio* menjadi parameter paling stabil dalam 48 simulasi penelitian dengan rata-rata akurasi tertinggi. Tidak semua data masukan digunakan oleh CTA untuk menentukan suatu kelas penggunaan lahan, dari 34 data masukan pada akurasi terbaik skema Level I terbaik membutuhkan 14 data masukan, sedangkan skema Level II terbaik membutuhkan 17 data masukan. Hal ini dikarenakan ketidakmurnian (*impurity*) data tersebut untuk dijadikan *node* pemisah pohon keputusan.
4. Hipotesis kedua dalam penelitian ini terbukti, dimana akurasi keseluruhan penelitian lebih dari 90%, dengan akurasi terbaik pada Kabupaten Purworejo sebesar 96,71% dan akurasi terbaik pada Kabupaten Kebumen sebesar 93,27%.

UCAPAN TERIMA KASIH

Ucapan terima kasih dan penghargaan disampaikan kepada Bapak Prof. Dr. Hartono, DEA., DESS. dan Bapak Drs. Projo Danoedoro, M.Sc., Ph.D. atas segala saran dan arahnya.

DAFTAR PUSTAKA

- Clark Labs. (2008). *Classification Tree Analysis*. Worcester. Clark University.
- Direktorat Pemetaan Tematik. (2012). *Norma, Standar, Prosedur, dan Kriteria - Survei dan Pemetaan Tematik Pertanahan*. Jakarta: Direktorat Pemetaan Tematik - Deputi Bidang Survei, Pengukuran, dan Pemetaan BPN-RI.
- Indrawati, L. (2009). *Klasifikasi Pohon Keputusan untuk Kajian Perubahan Penggunaan Lahan Kota*

- Semarang Menggunakan Citra Landsat TM/ETM+. *Tesis*. Yogyakarta. Universitas Gadjah Mada.
- Li, D., Di, K., & Li, D. (2000). Land Use Classification of Remote Sensing Image with GIS Data Based on Spatial Data Mining Techniques. *International Archives Of Photogrammetry And Remote Sensing*, XXXIII, 238-245.
- McCoy, R.M. (2005). *Field Methods in Remote Sensing*. New York. The Guilford Press.
- Saleh, F. (2014). Studi Komparasi Metode Fuzzy ARTMAP dan Multilayer Perceptron Untuk Klasifikasi Penggunaan Lahan Menggunakan Citra ALOS AVNIR-2. *Tesis*. Universitas Gadjah Mada. Yogyakarta.
- Samudra, I.S. (2007). Kajian Kemampuan Metode Jaringan Syaraf Tiruan untuk Klasifikasi Penutup Lahan dengan Menggunakan Citra Aster. *Tesis*. Yogyakarta. Universitas Gadjah Mada.
- Supribadi, K. (2014). Analisis Metode Support Vector Machine (SVM) untuk Klasifikasi Penggunaan Lahan Berbasis Penutup Lahan pada Citra ALOS AVNIR-2. *Tesis*. Yogyakarta. Universitas Gadjah Mada.
- Teddlie, C., & Yu, F. (2007). Mixed Methods Sampling: A Typology With Examples. *Journal of Mixed Methods Research*. 1, 77-100.
- Wu, X., & Kumar, V. (2009). *The Top Ten Algorithms in Data Mining*. Boca Raton. CRC Press.
- Zambon, M., Lawrence, R., Bunn, A., & Powell, S. (2006). Effect of Alternative Splitting Rules on Image Processing Using Classification Tree Analysis. *Photogrammetric Engineering and Remote Sensing*. 72, 25-30.