



MODEL KLASIFIKASI SPASIAL TINGKAT KERAWANAN KEBAKARAN HUTAN DAN LAHAN MENGGUNAKAN ALGORITME POHON KEPUTUSAN C5.0 DI KUBU RAYA

© Hak cipta milik IPB (Institut Pertanian Bogor)

FEBRI RAMADHAN



DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2017

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural U

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi berjudul Model Klasifikasi Spasial Tingkat Kerawanan Kebakaran Hutan dan Lahan menggunakan Algoritme Pohon Keputusan C5.0 di Kubu Raya adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

© Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2017

Febri Ramadhan
NIM G64130109

Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural University

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



ABSTRAK

FEBRI RAMADHAN. Model Klasifikasi Spasial Tingkat Kerawanan Kebakaran Hutan dan Lahan Menggunakan Algoritme Pohon Keputusan C5.0 di Kubu Raya. Dibimbing oleh MAYANDA MEGA SANTONI.

Kebakaran hutan dan lahan di Indonesia masih menjadi masalah yang sangat penting bagi perhutanan Indonesia. Salah satu kabupaten yang masih mengalami kebakaran hutan dan lahan adalah Kabupaten Kubu Raya, Kalimantan Barat. Perkembangan luas lahan perkebunan di kabupaten tersebut menjadi faktor utama terjadinya kebakaran hutan dan lahan. Untuk mengatasi lebih dini kebakaran hutan dan lahan tersebut, maka penelitian ini akan mengklasifikasikan tingkat kerawanan kebakaran hutan dan lahan menggunakan algoritme pohon keputusan C5.0 dengan 7 atribut yang berpotensi memengaruhi kebakaran hutan dan lahan. Ketujuh atribut tersebut adalah jarak dengan jalan raya, jarak dengan hutan lahan kering, jarak dengan pusat pemerintahan, jarak dengan pemukiman warga, jarak dengan perkebunan, jarak dengan aliran sungai, dan jarak dengan lahan gambut. Akurasi yang diperoleh pada penelitian ini adalah 90% dengan pembagian data uji dan data latih menggunakan *k-fold cross validation* dengan $K = 10$. Hasil akhir dari model tersebut divisualisasikan ke dalam bentuk peta kerawanan kebakaran hutan dan lahan. Kesimpulan penelitian ini adalah ketujuh atribut yang digunakan dapat menghasilkan *rules* dalam mengklasifikasikan tingkat kerawanan kebakaran hutan dan lahan dengan atribut yang berpengaruh adalah atribut jarak dengan perkebunan.

Kata Kunci: kebakaran hutan dan lahan, klasifikasi, Pohon Keputusan C5.0.

ABSTRACT

FEBRI RAMADHAN. Spatial Classification Model of Forest and Land Fire Vulnerability Level Using C5.0 Decision Tree Algorithm in Kubu Raya. Supervised by MAYANDA MEGA SANTONI.

Forest and land fires in Indonesia are still a very important issue for Indonesian forestry. One of the districts that are still experiencing forest and land fires is Kabupaten Kubu Raya, West Kalimantan. The increase in area that is used for plantations in the district is the main factor of forest and land fires. To address these forest and land fires, this study classified the extent of forest and land fires using the C5.0 decision tree algorithm with 7 attributes that have the potential to affect forest and land fires. The seven attributes are distance from highway, distance to dryland forest, distance from government center, distance to settlement, distance to plantation, distance to river flow, and distance to peatland. Accuracy obtained in this research is 90% with the distribution of test data and training data using k-fold cross validation with $K = 10$. The final result of the model is visualized into the form of a map with the forest and land fire vulnerability. The conclusion of this research is that the seven attributes used can generate rules in classifying the level of forest and land fire vulnerability with the distance to plantation as the most influencing attribute.

Keywords: classification, decision tree C5.0, land and forest fires.

- Hak Cipta Dilindungi Undang-Undang
© Hak Cipta milik IPB Institute of Pertanian Bogor
Bogor Agricultural University
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



MODEL KLASIFIKASI SPASIAL TINGKAT KERAWANAN KEBAKARAN HUTAN DAN LAHAN MENGGUNAKAN ALGORITME POHON KEPUTUSAN C5.0 DI KUBU RAYA

©

Hak cipta milik IPB (Institut Pertanian Bogor)

FEBRI RAMADHAN

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer
pada
Departemen Ilmu Komputer

**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2017**

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural U

Pengaji:

- 1 Dr Imas Sukaesih Sitanggang, SSi MKom
- 2 Lailan Sahrina Hasibuan, SKom MKom

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Judul Skripsi: Model Klasifikasi Spasial Tingkat Kerawanan Kebakaran Hutan dan Lahan menggunakan Algoritme Pohon Keputusan C5.0 di Kubu Raya

Nama : Febri Ramadhan
NIM : G64130109

Disetujui oleh

Mayanda Mega Santoni, SKomp MKom
Pembimbing

Diketahui oleh



Dr. Ir. Agus Bruno, MSi MKom
Ketua Departemen

Tanggal Lulus:

11 AUG 2017

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah *subhanahu wa ta'ala* atas segala karunia-Nya sehingga penulis dapat menyelesaikan penyusunan karya ilmiah dengan judul Model Klasifikasi Spasial Tingkat Kerawanan Kebakaran Hutan dan Lahan menggunakan Algoritme Pohon Keputusan C5.0 di Kubu Raya.

Penulis menyadari dalam penyusunan karya ilmiah ini masih banyak terdapat kesalahan dikarenakan masih kurangnya ilmu dan pengetahuan yang dimiliki oleh penulis. Oleh sebab itu, penulis mengucapkan terima kasih kepada pihak-pihak berikut yang telah mendukung dan membantu lancarnya proses penelitian hingga penyusunan laporan hasil penelitian tugas akhir ini. Terima kasih penulis sampaikan kepada:

- 1 Bapak Suhardi dan Ibu Nelviarti serta saudara kandung penulis atas doa, motivasi, dan dukungan baik secara moril dan materil yang telah diberikan hingga penulis bisa mencapai tahap sekarang ini.
- 2 Ibu Mayanda Mega Santoni, SKomp MKom selaku dosen pembimbing yang telah memberikan bimbingan dan arahannya sehingga penulis bisa menyelesaikan penulisan skripsi ini dengan baik.
- 3 Bapak Dr Ir Agus Buono, MSi MKom selaku Ketua Program Studi Ilmu Komputer IPB.
- 4 Ibu Dr Imas Sukaesih Sitanggang, SSi MKom dan Ibu Lailan Sahrina Hasibuan, SKom MKom selaku penguji atas saran dan masukan yang diberikan.
- 5 Seluruh dosen dan staf pegawai tata usaha Departemen Ilmu Komputer IPB yang telah memberikan bantuan dalam kelancaran dari perkuliahan hingga sampai ke tahap ini.
- 6 Seluruh sahabat penulis baik di Departemen Ilmu Komputer dan departemen lainnya.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan.

Bogor, Agustus 2017

Febri Ramadhan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



DAFTAR ISI

DAFTAR TABEL	vi
DAFTAR GAMBAR	vi
DAFTAR LAMPIRAN	vi
PENDAHULUAN	1
Latar Belakang	1
Perumusan Masalah	2
Tujuan Penelitian	2
Manfaat Penelitian	2
Ruang Lingkup Penelitian	3
METODE	3
Data Penelitian	3
Tahapan Penelitian	3
Pembentukan Data Set	3
Praproses Data	5
Pembagian Data	5
Pemodelan Pohon Keputusan	6
Pengujian Model Klasifikasi	7
Visualisasi Hasil	7
Lingkungan Penelitian	8
HASIL DAN PEMBAHASAN	8
Pembentukan Data Set	8
Iterasi Pertama	12
Iterasi Kedua	15
Visualisasi Hasil	18
SIMPULAN DAN SARAN	19
Simpulan	19
Saran	20
DAFTAR PUSTAKA	20
LAMPIRAN	22
RIWAYAT HIDUP	37

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



DAFTAR TABEL

1	Hasil pencarian tingkat kepadatan beberapa <i>grid</i>	11
2	Beberapa contoh hasil dari pencarian jarak terpendek antara hutan lahan kering dan titik tengah <i>grid</i>	12
3	Beberapa data sebelum dikategorikan	12
4	Pembagian kategori atribut jalan raya	12
5	Contoh data yang sesudah dikategorikan	13
6	Nilai evaluasi pada <i>fold</i> 6	14
7	Banyaknya kategori pada setiap atribut	16
8	Distribusi data uji dan data latih per- <i>fold</i>	16
9	Akurasi pada setiap <i>fold</i>	17
10	Nilai evaluasi pada <i>fold</i> 10	17
11	<i>Confusion matrix</i> model ke 10	18

DAFTAR GAMBAR

1	Tahapan Penelitian	4
2	Tahap pemotongan peta. (a) Peta Kubu Raya, (b) Potongan peta hutan lahan kering area Kalimantan Barat, (c) Penggabungan peta, dan (d) Hasil pemotongan	9
3	Proses pembuatan <i>grid</i> . (a) Peta Kubu Raya, (b) Peta Kubu Raya setelah ditambahkan <i>grid</i> , (c) Titik panas pada area <i>grid</i> , dan (d) Perbesaran beberapa <i>grid</i> yang terdapat titik panas	10
4	Kueri untuk mendapatkan data tingkat kepadatan	11
5	Kueri untuk mencari jarak terpendek antara titik tengah <i>grid</i> dengan hutan lahan kering	11
6	Grafik distribusi data perkelas	13
7	Grafik distribusi data kelas setelah <i>oversampling</i>	15
8	Peta tingkat kerawanan kebakaran hutan dan lahan Kabupaten Kubu Raya	19

DAFTAR LAMPIRAN

1	Data spasial yang digunakan	22
2	Kueri PostgreSQL untuk mendapatkan data jarak terpendek masing-masing atribut	26
3	Tabel selang jarak masing-masing atribut	27
4	Baris kode R untuk memperoleh model beserta informasi lainnya	28
5	Contoh perhitungan <i>information gain</i>	29
6	<i>Confusion matrix</i> data uji sebelum di- <i>oversampling</i>	31
7	Baris kode R untuk <i>k-fold cross validation</i>	32
8	Pohon keputusan model 10 (model terbaik)	33
9	<i>Confusion matrix</i> data uji setelah di- <i>oversampling</i>	36

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



PENDAHULUAN

Latar Belakang

Kebakaran hutan dan lahan di Indonesia masih terjadi hingga saat ini. Dampak dari kebakaran hutan dan lahan di Indonesia menimbulkan permasalahan-permasalahan lainnya, baik dari dampak dalam jangka pendek atau dalam jangka panjang. Dilihat dari kondisi geografisnya, Indonesia merupakan negara yang memiliki hutan tropis yang dominan dibandingkan dengan negara-negara lain. Keunggulan ini menyebabkan pentingnya dalam menjaga keutuhan hutan di Indonesia (Jaya *et al.* 2007).

Salah satu kabupaten di Indonesia yang mengalami kebakaran hutan dan tercatat oleh BNPB (Badan Nasional Penanggulangan Bencana) memiliki banyak titik api dari tahun 2000 sampai tahun 2015 adalah Kabupaten Kubu Raya Kalimantan Barat. Ditinjau dari potensi ekonomi, letak geografis, serta sumber daya alam yang unggul membuat wilayah ini menjadi target penting dalam pencegahan kebakaran hutan dan lahan di Indonesia (Hajeri *et al.* 2015).

Pencegahan terjadinya kebakaran hutan dan lahan sudah dilakukan sejak tahun 1982. Upaya pencegahan dilakukan dengan memprediksi terjadinya kebakaran hutan dan lahan pra kebakaran hingga penanggulangan kebakaran hutan dan lahan di saat atau pasca kebakaran. Hingga sekarang kebakaran hutan dan lahan di Indonesia masih menimbulkan kerugian yang sangat besar dalam sisi ekonomi, akomodasi, dan kesehatan (Jaya *et al.* 2007).

Berbagai penelitian telah dilakukan, dari meneliti timbulnya titik panas hingga meneliti kebakaran hutan dan lahan gambut sebagai aspek bahan bakar kebakaran hutan. Lahan dan hutan gambut merupakan penumpukan hasil dekomposisi yang tidak sempurna dari tumbuhan atau pepohonan yang tergenang di air. Tebalnya gambut bisa menjadi sumber bahan bakar dari kebakaran hutan dan lahan (Mallynur 2011).

Penelitian Hadi (2006) melakukan pembuatan model spasial tingkat kerawanan kebakaran lahan gambut di Kabupaten Bengkalis Provinsi Riau. Penelitian ini menyimpulkan bahwa kebakaran lahan gambut di Kabupaten Bengkalis Provinsi Riau dipengaruhi oleh keberadaan jaringan jalan, tipe tutupan lahan, ketebalan gambut dan tingkat kehijauan vegetasi. Aspek infrastruktur juga memiliki peran yang lebih besar dibanding faktor lingkungan fisik dalam penyebab terjadinya kebakaran lahan gambut.

Penelitian Samsuri (2008) telah membangun model spasial tingkat kerawanan kebakaran hutan dan lahan menggunakan metode *composite mapping analysis* (CMA). Penelitian yang dilakukan melibatkan 8 faktor yang dapat memengaruhi kebakaran hutan dan lahan yaitu tutupan lahan, jarak titik panas terhadap sungai, jarak titik panas terhadap jalan, jarak titik panas terhadap pusat desa, jarak titik panas terhadap pusat kota, penggunaan lahan, tipe tanah, dan sistem lahan. Penelitian ini menghasilkan empat faktor utama yakni tipe sistem lahan, tipe tutupan lahan, tipe tanah, dan fungsi kawasan yang berpengaruh terhadap kejadian kebakaran hutan dan lahan. Nilai akurasi pada penelitian tersebut ialah sebesar 66.76% untuk pengelompokan ke dalam tiga kelas dan 52.6% untuk pengelompokan ke dalam lima kelas. Besarnya curah hujan berpengaruh terhadap



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

jumlah titik panas dengan nilai koefisien determinasi sebesar 66.7 %. Faktor tingkat kepadatan titik panas memiliki pengaruh yang kecil terhadap tingkat kerusakan.

Selain teknik CMA, teknik *data mining* juga dapat digunakan untuk menyelesaikan permasalahan klasifikasi. Salah satu teknik *data mining* adalah algoritme pohon keputusan C5.0. Algoritme pohon keputusan C5.0 merupakan penyempurnaan dari algoritme pohon keputusan ID3 dan pohon keputusan C4.5 yang dibentuk oleh Ross Quinlan tahun 1987. Algoritme C5.0 telah dirancang untuk dapat menganalisis basis data substansial yang berisi ratusan atau lebih *record* serta variabelnya hingga ratusan *field* numerik atau nominal. Hasil klasifikasi algoritme C5.0 disajikan dalam bentuk sekumpulan aturan *if-then* atau berbentuk pohon keputusan agar lebih mudah dibaca oleh pengguna (Munawaroh *et al.* 2013).

Pada penelitian ini dibangun model klasifikasi spasial tingkat kerawanan kebakaran hutan dan lahan wilayah Kabupaten Kubu Raya menggunakan algoritme pohon keputusan C5.0. Faktor-faktor yang digunakan dari penelitian ini disesuaikan dengan faktor-faktor dari penelitian Samsuri (2008) yang diperoleh dalam bentuk data spasial.

Perumusan Masalah

Rumusan masalah penelitian ini adalah bagaimana mengklasifikasikan data spasial dalam menentukan tingkat kerawanan kebakaran hutan dan lahan di Kabupaten Kubu Raya dengan menggunakan algoritme pohon keputusan C5.0. Selain itu, dengan faktor-faktor yang telah dikumpulkan apakah bisa mengklasifikasikan tingkat kerawanan kebakaran hutan dan lahan di Kabupaten Kubu Raya.

Tujuan Penelitian

Tujuan dari penelitian ini adalah:

- 1 Menerapkan algoritme pohon keputusan C5.0 pada pengklasifikasian model spasial tingkat kerawanan kebakaran hutan dan lahan di Kabupaten Kubu Raya.
- 2 Mengevaluasi akurasi serta menerapkan teknik-teknik *data mining* pada pemodelan dalam mengidentifikasi tingkat kerawanan kebakaran hutan dan lahan di Kabupaten Kubu Raya.
- 3 Membuat peta kerawanan di wilayah Kabupaten Kubu Raya sebagai hasil akhir dari pengklasifikasian tingkat kerawanan kebakaran hutan dan lahan berdasarkan faktor-faktor yang diteliti.

Manfaat Penelitian

Manfaat dari penelitian ini diharapkan model spasial dalam bentuk peta kerawanan kebakaran hutan dan lahan tersebut dapat digunakan untuk pencegahan kebakaran hutan dan lahan. Oleh karena itu, model ini nanti akan dijadikan sebagai



pertimbangan oleh pihak terkait sebagai sebuah sistem dalam langkah awal mengatasi kebakaran hutan.

Ruang Lingkup Penelitian

Ruang lingkup dari penelitian ini adalah teknik klasifikasi menggunakan algoritme pohon keputusan C5.0. Selain itu, data yang digunakan dirujuk dari penelitian Samsuri (2008) dalam memilih atribut yang digunakan sesuai dengan topografi wilayah Kabupaten Kubu Raya.



Hak cipta milik IPB (Institut Pertanian Bogor)

METODE

Data Penelitian

Data yang digunakan pada penelitian ini diperoleh dari berbagai sumber. Data titik panas se-Indonesia diperoleh dari *Hotspot MCD14ML National Aeronautics and Space Administration* (NASA) yang diunduh pada tanggal 16 November 2016. Data titik panas diambil dari tanggal 2 Januari 2006 sampai dengan 31 Desember 2015.

Peta hutan lahan kering se-Indonesia, peta pemukiman penduduk se-Indonesia, peta perkebunan se-Indonesia, peta pemerintahan se-Indonesia, dan peta sungai se-Indonesia berasal dari Badan Informasi Geospasial (BIG) diunduh pada tanggal 7 Februari 2017. Peta jalan raya se-Indonesia diperoleh dari Geofabrik diunduh pada tanggal 7 Februari 2017 dan peta gambut se-Indonesia diperoleh dari *Wetlands International Indonesia* diunduh pada tanggal 26 Maret 2017.

Tahapan Penelitian

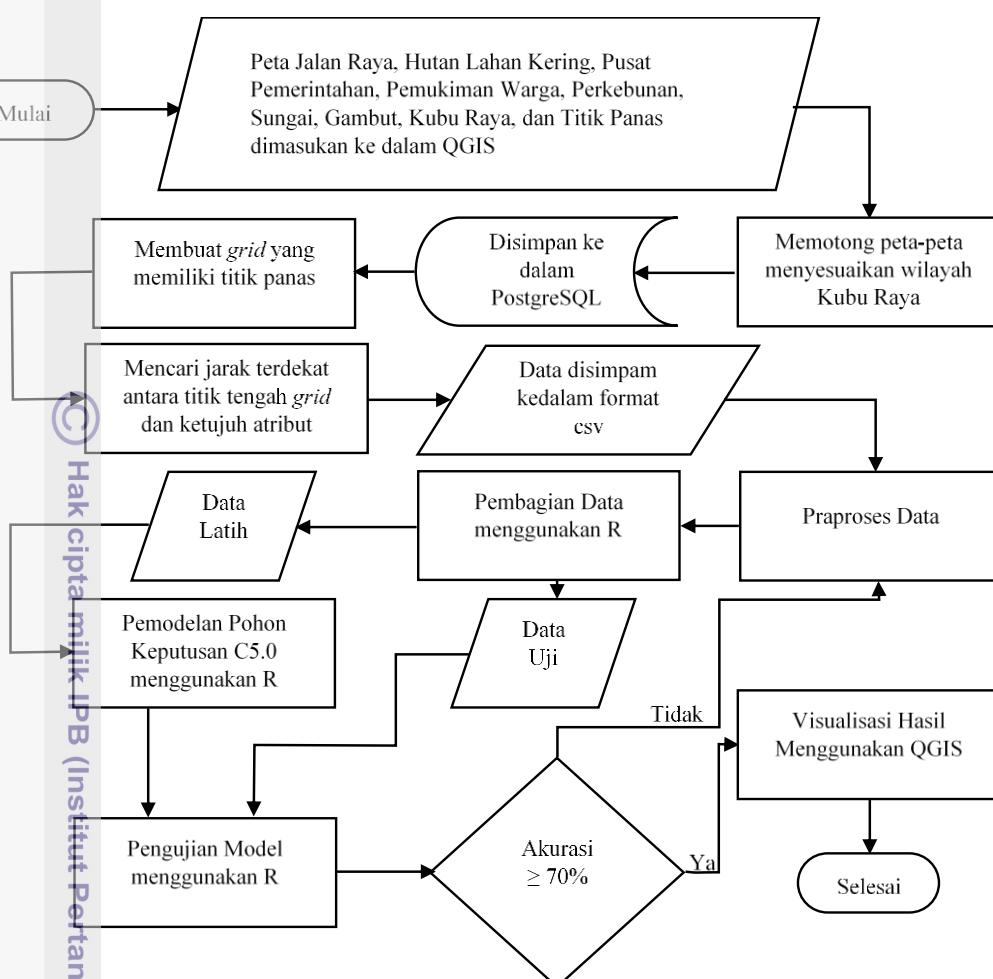
Tahapan-tahapan yang dilakukan pada penelitian ini secara garis besar terdiri dari pembentukan *data set*, praproses data, pembagian data, pemodelan pohon keputusan, pengujian model, dan visualisasi. Alur dari tahapan-tahapan tersebut dapat dilihat pada Gambar 1.

Pembentukan Data Set

Tahapan awal dilakukan operasi spasial untuk menyeleksi data titik panas, peta jalan raya, peta hutan lahan kering, peta pusat pemerintahan, peta pemukiman warga, peta perkebunan, peta sungai, dan peta gambut yang berada pada wilayah Kabupaten Kubu Raya. Penyeleksian dilakukan dengan memotong peta-peta dan titik panas terhadap peta batas wilayah Kubu Raya dengan menggunakan *tool Clip* pada QuantumGIS. Peta Kubu Raya diperoleh dari BIG yang digunakan sebagai acuan pemotongan area atribut dan titik panas.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Bogor Agricultural University



Gambar 1 Tahapan penelitian

Tahapan berikutnya semua peta dan titik panas dimasukkan kedalam DBMS PostgreSQL dengan bantuan *tool* PostGIS yang dimiliki pgAdminIII. Di dalam PostgreSQL data peta dan titik panas disimpan dalam data tabular.

Data tabular tersebut menyimpan informasi lokasi geometri yang disimpan di salah satu kolom. Penyimpanan semua peta dan titik panas kedalam data tabular bertujuan agar dapat diolah pada tahap-tahap berikutnya.

Data titik panas digunakan untuk memperoleh kelas atau variabel tak bebas. Variabel kelas merupakan tingkat kerawanan kebakaran hutan yang berasal dari kepadatan titik panas pada suatu *grid* tertentu. Menurut pakar semakin tinggi kepadatan maka semakin tinggi juga tingkat kerawanan kebakaran hutan dan lahan tersebut. Dalam penelitian ini, tingkat kepadatan titik panas berasal dari pembagian jumlah titik panas dengan luas area 1 km x 1 km yang dirujuk dari penelitian Samsuri (2008). Luas area 1 km x 1 km direpresentasikan ke dalam bentuk persegi (*grid*).

Grid merupakan persegi yang menjadi luas area dalam menentukan kepadatan. Metode *grid* adalah cara untuk membatasi gerombolan objek berdasarkan kedekatan dengan tetangganya yang berada pada suatu *grid* sebagai unit terkecil (Zhao *et al.* 2011). Setelah *grid* terbentuk, perhitungan kepadatan titik panas dilakukan dengan membagi jumlah titik panas pada masing-masing area *grid*.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

dengan luas area masing-masing *grid* dengan bantuan PostgreSQL pada pgAdmin III. Luas area diperoleh menggunakan fungsi ST_Area pada *library* PostGIS.

Variabel bebas yang menjadi atribut diperoleh dari jarak terdekat titik tengah masing-masing *grid* dengan ketujuh faktor yang digunakan pada penelitian ini. Pencarian jarak terdekat dilakukan dengan menggunakan fungsi-fungsi yang ada pada PostGIS seperti ST_Distance untuk mencari jarak, dan ST_Centroid untuk mencari titik tengah *grid*.

Setelah kelas dan atribut-atribut terbentuk, kedua sub data tersebut digabungkan dengan menyesuaikan lokasi masing-masing *grid* ke dalam tabel *data set*. Lokasi *grid* tetap disimpan dalam tabel *data set* untuk digunakan pada tahap visualisasi peta kerawanan. Jumlah kolom pada tabel *data set* setelah dibentuk ada 9 yaitu lokasi *grid*, jarak terdekat dengan jalan raya, jarak terdekat dengan hutan lahan kering, jarak terdekat dengan pusat pemerintahan, jarak terdekat dengan pemukiman warga, jarak terdekat dengan perkebunan, jarak terdekat dengan aliran sungai, jarak terdekat dengan lahan gambut, dan tingkat kepadatan titik panas.

Praproses Data

Tahapan awal praproses adalah dengan mengubah data interval menjadi data ordinal. Perubahan tersebut untuk mendiskrit nilai-nilai pada atribut dan kelas serta bertujuan mempermudah dalam mendeklarasikan suatu jarak. Pengubahan ke dalam bentuk ordinal dilakukan dengan cara membagi interval ke dalam 4 atau 5 kategori. Keempat kategori tersebut diantaranya dekat, sedang, jauh, dan sangat jauh. Label pada 5 kategori sama seperti 4 kategori dan ditambah dengan kategori label berada. Suatu atribut dibagi ke dalam 5 kategori jika atribut tersebut memiliki nilai minimum jarak terpendek dengan titik tengah *grid* sama dengan nol. Perbaahan data interval menjadi data ordinal pada variabel kelas diubah ke dalam 3 kategori. Tiga kategori tersebut adalah rendah, sedang, dan tinggi.

Tahap praproses lainnya adalah menganalisis statistik deskripsi dari data tersebut untuk bisa diputuskan akan dilakukan tindakan praproses yang seperti apa. Proses ini dilakukan untuk mempersiapkan data agar dapat digunakan oleh algoritme pohon keputusan C5.0 dengan menghasilkan akurasi yang baik.

Pembagian Data

Pada tahapan pembagian data, data dibagi menjadi dua, yaitu data uji dan data latih. Dalam penelitian ini, metode yang digunakan dalam proses pemisahan data latih dan data uji adalah metode *k-fold cross validation* dengan nilai $K=10$. *K-fold cross validation* merupakan metode yang membagi himpunan contoh secara acak menjadi K himpunan bagian. Menurut Han *et al.* (2012) *K-fold cross validation* dengan $K = 10$ merupakan pembagian himpunan yang baik dalam penentuan nilai K .

Penggunaan *cross validation* bertujuan untuk meningkatkan performa data keseluruhan. Performa data keseluruhan dinilai dari apakah data latih dan data uji tersebut mewakili seluruh *data set*. Selain meningkatkan performa data keseluruhan,



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

penggunaan *cross validation* diharapakan bisa meningkatkan akurasi model dengan jumlah data yang sedikit.

Tahapan awal pembagian data adalah dengan membagi data menjadi beberapa tabel sesuai dengan kategori kelas. Pemecahan tersebut bertujuan untuk mempermudah penggunaan secara seimbang data latih dan data uji pada setiap kelas. Jika menggunakan $K = 10$, maka pengambilan sub-data uji dilakukan dengan mengambil sub-data uji pada kelas rendah sebanyak 10%, kelas sedang 10%, dan kelas tinggi 10%. Setelah diperoleh sub-data uji dari ketiga jenis kategori, ketiga sub-data uji tersebut disatukan dan dijadikan data uji. Sisa semua dari data uji tersebut dijadikan data latih. Perlakuan tersebut dilakukan untuk semua data secara bergiliran. Oleh karena itu, semua data diperlakukan adil sehingga memiliki kesempatan yang sama menjadi data uji dan data latih.

Pemodelan Pohon Keputusan

Pemodelan model yang dilakukan menggunakan model klasifikasi pohon keputusan algoritme C5.0 dengan menggunakan data latih yang telah dibentuk. Algoritme C5.0 adalah pengembangan dari algoritme C4.5 dan juga ID3 (Patil *et al.* 2012). Algoritme C5.0 lebih baik daripada C4.5 dalam hal akurasi, kecepatan dan memori (Rulequest 2012). Model klasifikasi yang dibentuk oleh pohon keputusan algoritme C5.0 ini berbasis aturan terdiri dari kondisi *if-then (rule based)* dan direpresentasikan kedalam bentuk pohon keputusan. Pohon keputusan merupakan suatu diagram yang mirip dengan struktur *flowchart*, dimana masing-masing simpul merupakan atribut, masing-masing cabang menunjukkan nilai dari atribut, dan masing-masing simpul daun menunjukkan label kelas. Algoritme C5.0 melakukan pemilihan atribut yang akan diproses menggunakan *information gain* paling besar. Atribut yang memiliki *information gain* terbesar dipilih sebagai *parent* atau untuk *node* selanjutnya. Persamaan *entropy* dan *information gain* yang digunakan pada algoritme C5.0 adalah sebagai berikut (Han *et al.* 2012):

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

$Info(D)$ adalah nilai *entropy* dari sampel data D , m adalah jumlah kelas yang ada di atribut, sedangkan p_i adalah peluang dari kelas i atau rasio dari kelas. Nilai entropy yang dihasilkan untuk mengklasifikasi *tuple* dari D berdasarkan partisi oleh A menggunakan formula pada Persamaan 2 (Han *et al.* 2012):

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

Partisi *tuple* di D pada beberapa atribut A memiliki nilai v yang berbeda $\{a1, a2, ..., av\}$ dari data latih. Atribut A digunakan untuk memisahkan D ke dalam v partisi atau sub himpunan $\{D1, D2, ..., Dv\}$. $\frac{|D_j|}{D}$ merupakan bobot partisi ke- j .

Information gain yang diperoleh pada atribut A menyatakan bahwa ada berapa banyak cabang yang akan diperoleh pada A . Atribut A dengan *information gain* tertinggi dipilih sebagai atribut pada node. Untuk mendapatkan nilai *gain*, dapat menggunakan Persamaan 3 (Han *et al.* 2012).

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

Pembentukan model klasifikasi tersebut dilakukan menggunakan *package* yang tersedia pada aplikasi R yaitu C5.0. Fungsi yang digunakan adalah fungsi c5.0 dengan menyertakan data latih yang digunakan.

Pengujian Model Klasifikasi

Setelah mendapatkan model, tahapan selanjutnya adalah melakukan pengujian terhadap model yang diperoleh. Dalam pengujian model, data uji dari *data set* digunakan sebagai data masukan. Pada proses pengujian kelas pada data uji tersebut, kelas yang diprediksi dibandingkan dengan data aktualnya. Prediksi kelas dilakukan dengan bantuan perangkat lunak R menggunakan fungsi Predict pada *package* C5.0.

Untuk mengukur seberapa baik hasil pengujian model ini, maka digunakan metode *confusion matrix* dengan menggunakan *package* Caret. *Confusion matrix* merupakan metode untuk melihat akurasi yang merepresentasikan data mana saja yang tidak sesuai dengan label aslinya. Akurasi merupakan sebuah ukuran evaluasi dalam mengetahui keakuratan model yg dibuat pada saat pengujian.

Hasil pemetaan *confusion matrix* digunakan untuk mencari persentase seberapa banyak model bisa melabelkan kelas sesuai dengan kelas *real* pada data uji. Persentase tersebut merupakan perwakilan rataan nilai kebenaran model dalam melabelkan kelas *data set* tersebut. Akurasi yang dicapai menentukan evaluasi yang akan dilakukan untuk tahapan selanjutnya. Berikut pada Persamaan 4 adalah perhitungan nilai akurasi:

$$\text{Akurasi} = \frac{\text{Jumlah data uji terklasifikasikan benar}}{\text{Jumlah data uji keseluruhan}} \times 100\% \quad (4)$$

Selain menghitung akurasi, pengujian model klasifikasi juga menghitung *precision* dan *recall*. Menurut Han *et al* (2012) *precision* adalah tingkat ketepatan antara data yang berikan oleh suatu model dengan data yang diminta oleh pengguna sedangkan *recall* adalah tingkat keberhasilan suatu model dalam menentukan prediksi yang relevan. Persamaan 5 menunjukkan perhitungan *precision* dan persamaan 6 menunjukkan perhitungan *recall*.

$$Precision = \frac{True Positive}{True Positive + False Positive} \quad (5)$$

$$Recall = \frac{True Positive}{True Positive + False Negative} \quad (6)$$

Keterangan:

- True Positive* = Data dengan nilai aktual benar dan nilai prediksi benar
- False Positive* = Data dengan nilai aktual tidak benar dan nilai prediksi benar
- False Negative* = Data dengan nilai aktual benar dan nilai prediksi tidak benar

Visualisasi Hasil

Visualisasi pada penelitian ini merupakan pemetaan kembali *grid* yang telah diklasifikasikan kelas tingkat kerawannya berdasarkan ketujuh atribut. Data yang

- Hak Cipta Dilindungi Undang-Undang
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

digunakan adalah semua *data set*. *Data set* tersebut diklasifikasikan pada model yang terbentuk untuk mendapatkan kelas kerawannya. Setelah mendapatkan kelas kerawanan, *data set* yang memiliki atribut lokasi tersebut dipetakan kembali menggunakan perangkat lunak QuantumGIS. Pengolahan pada QuantumGIS antara lain pewarnaan tingkat kerawanan, pelabelan kelas, pemberian koordinat dalam bentuk *grid*, pemberian arah mata angin, pemberian *scalebar*, dan pemberian legenda.

Peta tersebut diharapkan menampilkan informasi tingkat kerawanan kebakaran hutan dan lahan dengan jelas. Peta kerawanan dibagi ke dalam beberapa kategori, diantaranya rawan kebakaran tinggi, rawan kebakaran sedang, dan rawan kebakaran rendah. Setiap kategori diberi warna yang berbeda seperti warna kuning untuk tingkat kerawanan rendah, warna jingga untuk tingkat kerawanan sedang, dan warna merah untuk tingkat kerawanan tinggi.

Lingkungan Penelitian

Perangkat lunak yang digunakan dalam penelitian ini adalah:

- Sistem operasi Windows 10 Home Edition 64-bit sebagai sistem operasi
- Bahasa pemrograman R digunakan dalam membuat model pohon keputusan
- RStudio versi 0.99.902 sebagai *graphical user interface* bahasa R
- Microsoft Excel 2016 untuk membaca dan mengolah data dengan format csv
- Quantum GIS 2.18.0 untuk melihat plot data spasial dan analisis data spasial
- PostgreSQL 9.4 sebagai sistem manajemen basis data.

Perangkat keras yang digunakan dalam penelitian ini adalah komputer personal dengan spesifikasi:

- Processor Intel Core 2 Duo
- RAM 4GB
- Harddisk 250 GB HDD
- VGA Intel(R) 4 Series Express Chipset Family.

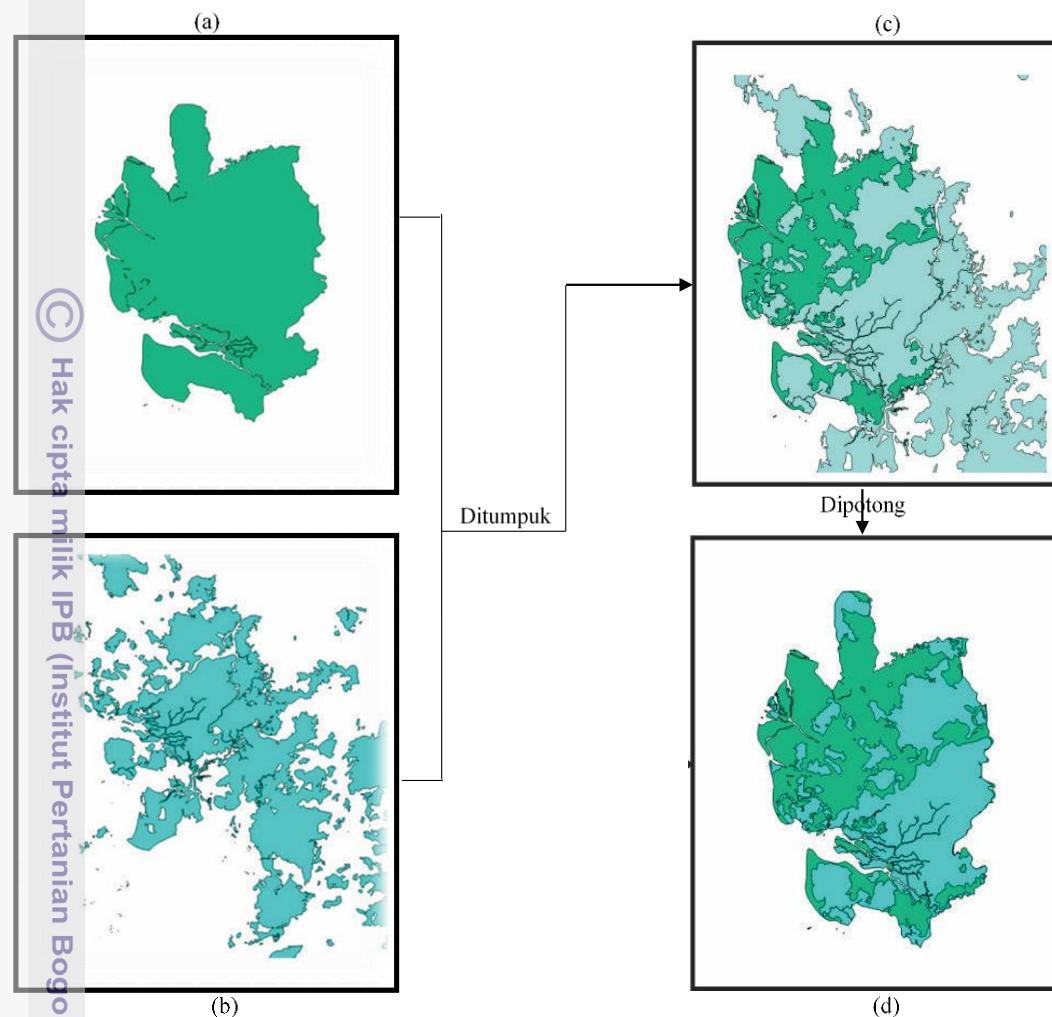
HASIL DAN PEMBAHASAN

Pembentukan *Data Set*

Tahapan awal pembentukan *data set* adalah pemindahan data spasial (peta) ke PostgreSQL. Sebelum dipindahkan, peta terlebih dahulu dipotong sesuai dengan bentuk geometri wilayah Kabupaten Kubu Raya menggunakan Quantum GIS. Quantum GIS digunakan untuk melakukan pengolahan secara geometri, sedangkan pengolahan data tabular menggunakan DBMS PostgreSQL.

Data spasial yang digunakan sebagai atribut antara lain, titik panas Kubu Raya, peta jalan raya, peta hutan lahan kering, peta pusat pemerintahan, peta pemukiman warga, peta perkebunan, peta aliran sungai, dan peta sebaran lahan

gambut yang dapat dilihat pada Lampiran 1. Sebelum diubah ke dalam bentuk data tabular, peta tersebut dipotong terlebih dahulu.



Gambar 2 Tahap pemotongan peta. (a) Peta Kubu Raya, (b) Potongan peta hutan lahan kering area Kalimantan Barat, (c) Penggabungan peta, dan (d) Hasil pemotongan

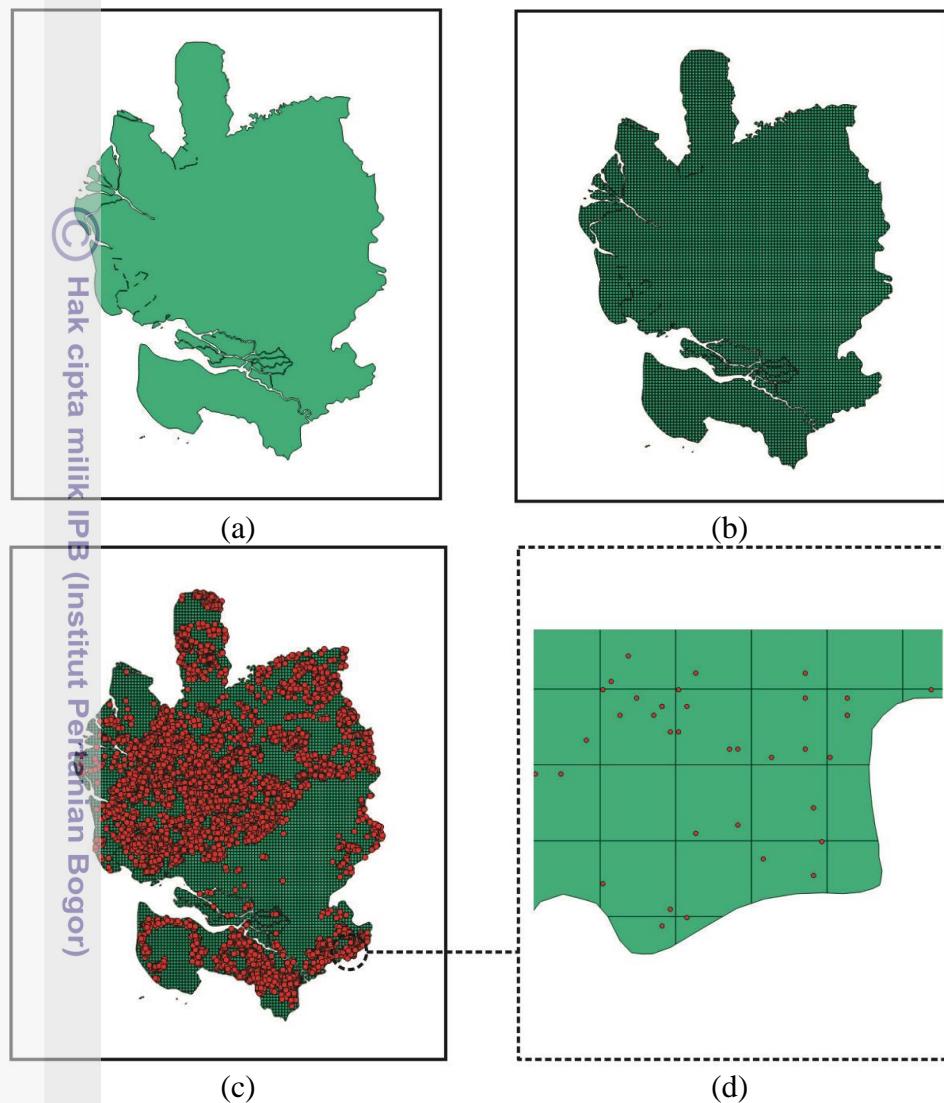
Pemotongan luas area semua peta atribut dilakukan dengan menggunakan *tool Clip* pada Quantum GIS. Setelah dipotong, semua peta dimasukan ke dalam DBMS pgAdmin III dengan bantuan *tool PostGIS* pada pgAdmin III untuk pengolahan secara tabular pada tahapan berikutnya. Gambar 2 merupakan contoh pemotongan luas hutan lahan kering se-Indonesia yang dipotong menjadi hutan lahan kering se-Kabupaten Kubu Raya.

Tahap selanjunya adalah membuat tingkat kepadatan titik panas (variabel tak bebas) dengan memerhatikan banyaknya jumlah titik panas pada suatu area tertentu. Dalam penelitian ini luas *grid* adalah 1 km x 1 km yang direpresentasikan kedalam bentuk *grid*. Pembentukan *grid* dilakukan dengan bantuan *tool* yang terdapat pada perangkat lunak Quantum GIS. Tahapan pembentukan *grid* hingga perbesaran beberapa *grid* yang memiliki titik panas dapat dilihat pada Gambar 3.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Gambar 3 Proses pembuatan *grid*. (a) Peta Kubu Raya, (b) Peta Kubu Raya setelah ditambahkan *grid*, (c) Titik panas pada area *grid*, dan (d) Perbesaran beberapa *grid* yang terdapat titik panas

Rumus mencari tingkat kepadatan adalah membagi jumlah titik panas yang terdapat pada area *grid* dengan luas area *grid* tersebut. *Grid* yang memiliki titik panas dijadikan objek kelas tingkat kepadatan titik panas. Total baris yang dihasilkan yaitu 2 609 baris. Kueri untuk mencari tingkat kepadatan titik panas serta sebagai tingkat kerawanan kebakaran hutan dan lahan dapat dilihat pada Gambar 4. Hasil dari kueri tingkat kepadatan beberapa *grid* dapat dilihat pada Tabel 1.

```
create table kelas as SELECT a.geom, (st_area(a.geom) * 1000000) /
80.6933971701611 as luas, count(b.geom) as jumlahTitikPanas, st_centroid(a.geom) as titikTengahGrid,
count(b.geom) / (st_area(a.geom) * 1000000) / 80.6933971701611) as tingkatkepadatan from grid a join hotspot b on
st_contains(a.geom,b.geom) group by a.geom
```

Gambar 4 Kueri untuk mendapatkan data tingkat kepadatan

Tabel 1 Hasil pencarian tingkat kepadatan beberapa *grid*

Grid	Luas	Jumlah Hotspot	Tingkat Kepadatan
MULTIPOLYGON(((109.070167954719 -0.277901330351,109.079150905079 -0.277901330351,109.079150905079 -0.286884280711,109.070167954719 -0.286884280711,109.070167954719 -0.277901330351)))	1	1	1
MULTIPOLYGON(((109.070167954719 -0.268918379991,109.079150905079 -0.268918379991,109.079150905079 -0.277901330351,109.070167954719 -0.277901330351,109.070167954719 -0.268918379991)))	1	8	8

Tahap berikutnya adalah pembentukan variabel bebas (atribut). Variabel bebas berasal dari jarak terdekat antara masing-masing faktor dengan titik tengah *grid*. Pencarian jarak terdekat dilakukan dengan menggunakan fungsi ST_Distance yang telah tersedia pada *library* PostGIS. Gambar 5 menunjukkan salah satu contoh kueri pencarian antara atribut hutan lahan kering dengan titik tengah *grid*. Kueri semua jarak terdekat antara titik tengah *grid* dengan semua atribut dapat dilihat pada Lampiran 2.

```
select st_distance(a.centroid, b.geom)*111.322 as jarak from
kelas a, jalanrayakuburaya b where centroid = kelas.centroid
order by jarak limit 1
```

Gambar 5 Kueri untuk mencari jarak terpendek antara titik tengah *grid* dengan hutan lahan kering

Keluaran jarak yang dihasilkan masih berbentuk satuan derajat. Oleh Karena itu, agar bisa direpresentasikan kedalam kilometer maka jarak tersebut dikonversi menjadi satuan kilometer dengan cara mengalikan dengan angka 111,322. Beberapa contoh hasil dari pencarian jarak terpendek antara hutan lahan kering dan titik tengah *grid* dari kueri pada Gambar 5 dapat dilihat pada Tabel 2.

Setelah mendapatkan jarak terpendek masing-masing faktor dan mendapatkan tingkat kepadatan, kedua data tersebut disatukan. Penyatuan dilakukan dengan menyesuaikan lokasi *grid* yang disimpan pada setiap tabel. Tabel 3 menunjukkan beberapa baris data setelah digabungkan.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Tabel 2 Beberapa contoh hasil dari pencarian jarak terpendek antara hutan lahan kering dan titik tengah grid

Lokasi Titik Tengah Grid	Jarak Terpendek (km)
"POINT(109.309 -0.295)"	8.40834425620369
"POINT(109.318 -0.296)"	9.330761576125
"POINT(109.268 -0.19)"	7.99334608742896

Tabel 3 Beberapa data sebelum dikategorikan

jarak jalan raya	jarak hutan lahan kering	jarak pemerintahan	jarak pemukiman	jarak perkebunan	jarak sungai	jarak gambut	Ke Padatan
16.575	8.408	9.286	0.452	2.815	7.586	14.216	1
25.851	9.331	9.715	1.252	2.784	6.655	14.367	1
24.217	7.993	10.803	3.148	2.723	4.877	14.852	2
41.614	8.855	12.852	0.336	2.631	2.832	16.019	1

Iterasi Pertama

Praproses Data

Pada tahapan praproses dilakukan pengelompokan nilai-nilai pada setiap atribut dan kelas. Pengelompokan dilakukan dengan melabelkan nilai-nilai numerik ke dalam beberapa kategori. Pengelompokan dilakukan berdasarkan interval-interval yang berasal dari pembagian antara selisih batas minimum dan batas maksimum dengan jumlah label. Penentuan jumlah label ditentukan berdasarkan 5 kategori diantaranya label berada, dekat, sedang, jauh, dan sangat jauh. Label berada hanya diperuntukan terhadap atribut yang memiliki jarak minimum sama dengan nol. Variabel kelas atau tingkat kepadatan dibagi ke dalam 3 kategori diantaranya rendah, sedang, dan tinggi. Tabel 4 menunjukkan contoh rentang dalam menentukan kategori jarak pada atribut jarak jalan raya dan untuk atribut lainnya dapat dilihat pada Lampiran 3.

Tabel 4 Pembagian kategori atribut jalan raya

Kategori	Rentang
Dekat	$X \leq 17.76215 \text{ km}$
Sedang	$17.76215 \text{ km} < X \leq 35.52429 \text{ km}$
Jauh	$35.52429 \text{ km} < X \leq 53.28644 \text{ km}$
Sangat Jauh	$> 53.28644 \text{ km}$

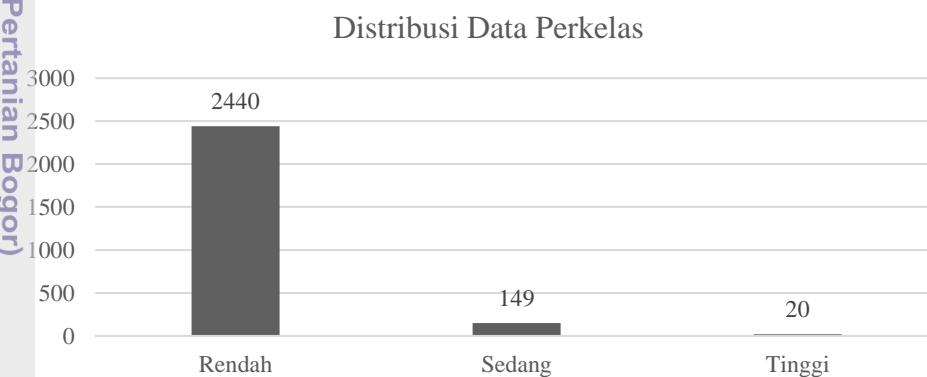
Tahap berikutnya adalah mengubah semua nilai data atribut dari numerik ke kategorik. Sebagai contoh, pada data atribut jarak dengan jalan raya memiliki nilai 15 km yang artinya jarak titik tengah grid ke jalan raya sebesar 15 km. Data ini diubah ke dalam kategori “dekat” yang telah ditentukan pada Tabel 4. Proses

perubahan semua data atribut dan kelas dari numerik ke kategorik dilakukan dengan menggunakan perangkat lunak Microsoft Excel. Contoh data yang telah dikategorikan untuk semua data atribut dan kelas dapat dilihat pada Tabel 5.

Tabel 5 Contoh data yang sesudah dikategorikan

jarak jalan raya	jarak hutan lahan kering	jarak pemerintahan	jarak pemukiman	jarak perkebunan	jarak sungai	jarak gambut	Ke Padatan
Dekat	Sedang	Dekat	Dekat	Berada	Dekat	Dekat	Rendah
Sedang	Berada	Sedang	Jauh	Dekat	Sangat Jauh	Berada	Rendah
Sedang	Berada	Dekat	Sangat Jauh	Jauh	Dekat	Dekat	Sedang
Jauh	Sedang	Sangat Jauh	Dekat	Berada	Sedang	Sedang	Tinggi
Dekat	Dekat	Sedang	Berada	Dekat	Dekat	Berada	Sedang

Distribusi kelas yang dihasilkan setelah pengelompokan diantaranya kelas rendah memiliki 2 440 baris, kelas sedang 149 baris, dan kelas tinggi 20 baris yang dapat dilihat pada Gambar 6. Data tersebut selanjutnya dimodelkan menggunakan perangkat lunak R dengan package C5.0.



Gambar 6 Grafik distribusi data perkelas

Pembagian Data

Pembagian data dilakukan dengan menggunakan metode *k-fold cross validation* dengan K = 10 menggunakan R. *Library* yang digunakan adalah *plyr* untuk mengolah *dataframe*. Tahapan awal data tersebut dilakukan pengacakan menggunakan fungsi acak yang terdapat pada R. Setelah dilakukan pengacakan, data tersebut dibagi ke dalam 3 *data frame* sesuai dengan jumlah kelas pada data penelitian ini. Setelah itu masing-masing ketiga *data frame* tersebut dilakukan pengacakan kembali dan diberi label indeks dari 1 sampai dengan 10 pada setiap baris data.

- Hak Cipta Dilindungi Undang-Undang
Hak cipta milik IPB Institut Pertanian Bogor
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB. 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Hak Cipta Dilindungi Undang-Undang

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Pemodelan Pohon Keputusan

Model klasifikasi dibuat menggunakan algoritme pohon keputusan C5.0 dengan menggunakan *library* C5.0 untuk pemodelan, dan Caret untuk mendapatkan *confusion matrix*, *precision* dan *recall*. Pemodelan dilakukan seiring dilakukannya pembagian data sebanyak 10 *fold*. Oleh karena itu pada pemodelan menggunakan algoritme C5.0 ini diperoleh 10 model yang mewakili setiap *fold*-nya. Kode program R yang digunakan dapat dilihat pada Lampiran 4.

Pemodelan dilakukan dengan menggunakan fungsi C5.0 dengan menyertakan tabel data latih dan data uji. Fungsi tersebut dijalankan sebanyak 10 kali sesuai jumlah *fold* pada penelitian ini. Tahap awal dimulai dengan menghitung *information gain* setiap. *Information gain* tersebut digunakan untuk menentukan *node-node* atau sebagai *parent*. Contoh perhitungan *information gain* beberapa data dapat dilihat pada Lampiran 5.

Pengujian Model Klasifikasi

Setelah diperoleh 10 model, masing-masing model dilakukan pengujian. Pengujian dilakukan dengan mengklasifikasikan kelas data uji. Kelas yang diprediksi dihitung persentasenya dari data yang sesuai dengan kelas *real* data uji. Prediksi yang dilakukan menggunakan fungsi Predict pada *library* C5.0. Perhitungan persentase tersebut dilakukan dengan menggunakan metode *confusion matrix*.

Confusion matrix diperoleh dengan menggunakan fungsi ConfusionMatrix yang terdapat pada *library* Caret. Kesepuluh model tersebut menghasilkan 10 akurasi diambil rataannya untuk dijadikan akurasi akhir. Rataan akurasi yang diperoleh adalah 93% dengan model tertinggi pada pada *fold* keenam. Lampiran 6 menunjukkan *confusion matrix* data uji ke 10 *fold*.

Pada model tersebut, kelas sedang dan kelas tinggi tidak terkласifikasi oleh model pada proses pengujian data uji. Hal tersebut terjadi karena tidak seimbangnya data dalam melakukan pemodelan. Terlihat pada Tabel 6 *precision* dan *recall* kelas sedang dan kelas tinggi memiliki nilai 0. Sebuah himpunan data dikatakan menjadi tidak seimbang jika terdapat satu kelas yang direpresentasikan dalam jumlah *instance* yang kecil bila dibandingkan dengan jumlah *instance* kelas yang lainnya.

Tabel 6 Nilai evaluasi pada *fold* 6

	Kelas		
	Rendah	Sedang	Tinggi
Precision	0,93	NaN	NaN
Recall	1,00	0	0

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

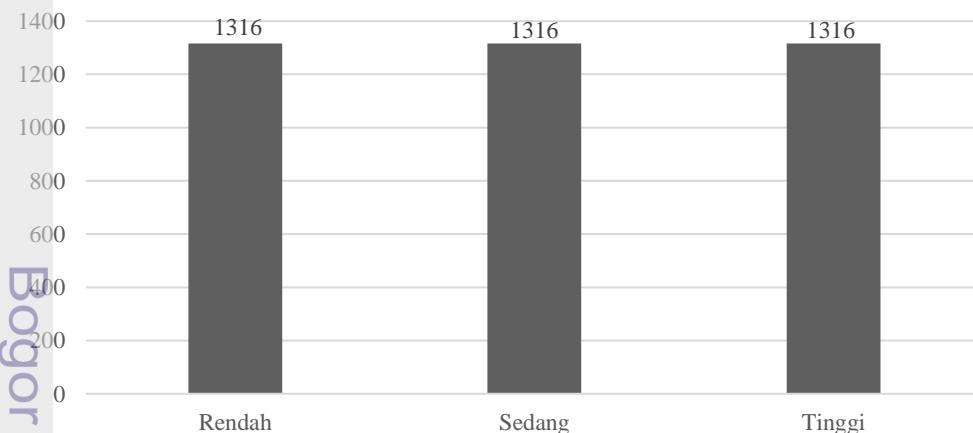
Praproses Data

Praproses iterasi dua dilakukan dengan menambah jumlah data dengan menggunakan teknik *oversampling acak*. Teknik *oversampling acak* adalah cara untuk menambah data kelas minor dengan membuat kombinasi lain antara kategori-kategori yang dimiliki kelas minor tersebut. Penggunaan *oversampling acak* memperhatikan sifat-sifat dan penyebaran populasi agar diperoleh sampel yang representatif (Margono 2004). Selain itu, penggunaan *oversampling* dilakukan dengan memperhatikan jumlah kombinasi fitur maksimal pada setiap kelas.

Kelas minor kategori sedang memiliki fitur maksimal 3 200 kombinasi data yang bisa menjadi *instance* tambahan dalam *oversampling* dan kelas minor kategori tinggi memiliki maksimum *oversampling* sebanyak 1 296 kombinasi baris. Kelas tinggi hanya bisa memiliki tambahan data maksimum sebanyak 1 296 baris ditambah dengan data *real* 20 baris menjadi 1316 baris.

Oleh karena itu, jumlah masing-masing kelas rendah dan kelas sedang mengikuti jumlah kelas tinggi yang hanya memiliki 1 316 baris. Oleh karena itu kelas rendah dipangkas secara acak menjadi 1 316, sedangkan kelas rendah dan tinggi di-*oversampling* menjadi 1 316 baris dengan total semua data menghasilkan 3 948 baris yang dapat dilihat pada Gambar 7.

Distribusi Perkelas Setelah Oversampling



Gambar 7 Grafik distribusi data kelas setelah *oversampling*

- Hak Cipta Dilindungi Undang-Undang**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Oversampling dilakukan menggunakan bantuan perangkat lunak Microsoft Excel. Penambahan data kombinasi baru dilakukan dengan memilih secara acak kategori-kategori yang dimiliki masing-masing atribut. Data baru tersebut diperbanyak hingga ribuan kali dengan harapan semua kombinasi kategori antara setiap atribut muncul. Setelah diperbanyak, data kombinasi baru tersebut disaring untuk menghilangkan duplikasi. Penyaringan duplikasi dilakukan dengan menggunakan bantuan fungsi *remove duplicate* pada Microsoft Excel. Hasil yang diperoleh adalah data baru yang berasal dari semua kombinasi baru pada setiap kelas.

Data *oversampling* kelas sedang sebanyak 3 200 data dipangkas sebanyak 2 033 baris secara acak. Pemangkasan tersebut dilakukan untuk menyesuaikan dengan kelas tinggi yang hanya memiliki 1 316 baris beserta data asli. Setelah dipangkas, hasil yang diperoleh adalah 1 167 baris data *oversampling*. Selanjutnya data tersebut disatukan dengan data asli sebanyak 149 baris dan menghasilkan total data 1 316 baris. Begitu juga dengan kelas tinggi dilakukan *oversampling* tanpa dilakukan pemangkasan. Tabel 7 menunjukkan jumlah kategori yang dimiliki setiap atribut pada masing-masing kelas.

Tabel 7 Banyaknya kategori pada setiap atribut

Kelas	Jalan Raya	Jarak Hutan Lahan Kering	Jarak Pusat Perintah an	Jarak Mukim an	Jarak Per kebun an	Jarak Sungai 3	Jarak Gambut 3	Maksimal Oversampling
Sedang	5	4	5	4	5	4	4	3200
Tinggi	3	3	2	2	3	3	3	1296

Pembagian Data

Pembagian data iterasi kedua sama seperti iterasi pertama yaitu dengan menggunakan metode *k-fold cross validation* dengan K = 10 menggunakan R. Pada setiap perulangan akan diambil kurang lebih 131 baris dari setiap kelas yang digabungkan menjadi satu set data uji. Sisa dari masing-masing ketiga *data frame* tersebut sekitar 1 185 digabungkan menjadi data latih. Setiap perulangan menggunakan kurang lebih 393 data uji dan 3 555. Data latih dan data uji tersebut dilatih menggunakan pohon keputusan C5.0 hingga diperoleh suatu model pada setiap perulangannya. Tabel 8 menunjukkan distribusi data untuk masing-masing *fold* dan untuk baris kode R dapat dilihat pada Lampiran 7.

Tabel 8 Distribusi data uji dan data latih per-*fold*

K fold	1	2	3	4	5	6	7	8	9	10
Data Uji	396	396	393	396	393	396	393	396	393	396
Data Latih	3552	3552	3555	3552	3555	3552	3555	3552	3555	3552

Pemodelan Pohon Keputusan

Model klasifikasi dibuat menggunakan algoritme pohon keputusan C5.0 sama seperti iterasi pertama. Pemodelan dilakukan seiring dilakukannya pembagian data sebanyak 10 *fold*. Oleh karena itu pada pemodelan menggunakan algoritme C5.0 ini diperoleh 10 model yang mewakili setiap *fold*-nya. Pemodelan dilakukan dengan menggunakan fungsi C5.0 dengan menyertakan tabel data latih dan data uji. Fungsi tersebut dijalankan sebanyak 10 kali sesuai jumlah *fold* pada penelitian ini.

Pengujian Model Klasifikasi

Pengujian model klasifikasi dilakukan dengan mengklasifikasikan kelas data uji sama seperti pengujian iterasi pertama. Kelas yang diprediksi akan dihitung persentase data yang sesuai dengan kelas *real* data uji. Perhitungan persentase tersebut dilakukan dengan menggunakan metode *confusion matrix*.

Kesepuluh model tersebut menghasilkan 10 akurasi yang diambil rataanya untuk dijadikan akurasi akhir. Rataan yang diperoleh adalah 90.55% dengan model tertinggi pada pada *fold* ketujuh. Tabel 9 menunjukkan hasil semua akurasi setiap *fold* dan Tabel 10 menunjukkan nilai *precision* dan *recall* pada *fold* kesepuluh. Bila dibandingkan dengan Tabel 6 terlihat nilai *precision* dan *recall* setiap kelas naik setelah dilakukan *oversampling*. Hal ini menunjukkan bahwa model bisa mengklasifikasikan setiap kelas dengan akurasi rataan 90%.

Tabel 9 Akurasi pada setiap *fold*

K-fold	1	2	3	4	5	6	7	8	9	10
Akurasi	87.63	89.14	92.62	88.89	92.88	92.17	93.89	90.91	87.02	90.40

Turunnya akurasi rataan setelah *oversampling* disebabkan oleh model sebelum *oversampling* tidak bisa mengklasifikasikan kelas sedang dan kelas tinggi. Selain itu, jumlah kelas rendah yang banyak dari kelas lainnya menyebabkan akurasi bergantung pada banyaknya jumlah kelas rendah dibanding kelas lainnya. Semakin banyak kelas rendah yang diuji dibandingkan kelas lainnya, maka semakin tinggi nilai akurasi yang dihasilkan. Oleh karena itu, nilai *recall* menurun pada kelas rendah yang dapat dilihat pada Tabel 10.

Tabel 10 Nilai evaluasi pada *fold* 10

	Kelas		
	Rendah	Sedang	Tinggi
<i>Precision</i>	0,94	0,87	0,95
<i>Recall</i>	0,89	0,97	0,91

Model yang digunakan sebagai model dalam menentukan tingkat kepadatan atau hasil akhir adalah model yang akurasinya mendekati rataan dengan atribut yang paling berpengaruh adalah atribut jarak dengan perkebunan. Model yang mendekati akurasi rataan adalah model ke 10 dapat dilihat *rule* dan persentase penggunaan atributnya pada Lampiran 8. Contoh *rule* yang dihasilkan pada model

ke 10 adalah “Jika jarak perkebunan dekat; jarak pemukiman dekat; jarak pemerintahan sangat jauh; berada pada hutan lahan kering; jarak dengan lahan gambut dekat; jarak sungai dekat; dan jarak jalan raya sangat jauh maka wilayah tersebut memiliki tingkat kepadatan titik panas yang tinggi”. Selanjutnya, *confusion matrix* pada model ke 10 dapat dilihat pada Tabel 11. *Confusion matrix* model lainnya dapat dilihat pada Lampiran 9.

Tabel 11 *Confusion matrix* model ke 10

Prediksi	Aktual		
	Rendah	Sedang	Tinggi
Rendah	128	13	10
Sedang	3	110	1
Tinggi	1	9	121

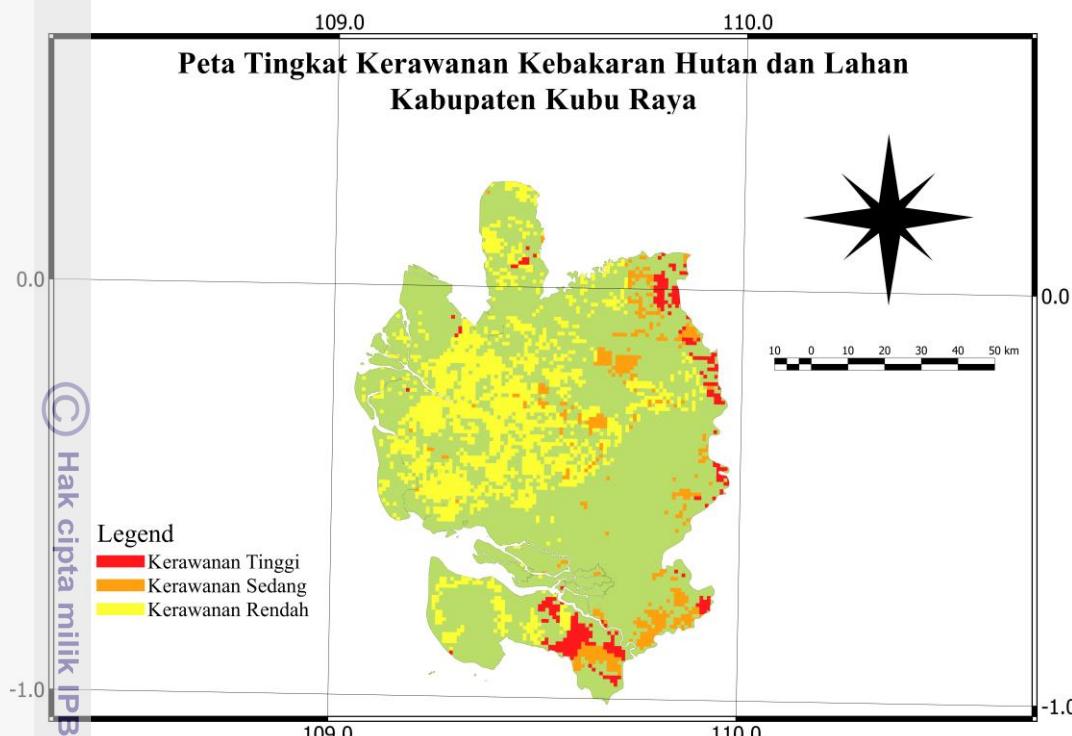
Visualisasi Hasil

Tahapan akhir dari penelitian ini adalah memvisualisasikan *grid* yang telah diklasifikasikan dari model pohon keputusan kedalam bentuk spasial (peta). Langkah awal visualisasi adalah mencari semua tingkat kepadatan sebagai kelas kerawanan masing-masing *grid* pada data awal atau data sebelum *oversampling*. Hasil dari pengklasifikasian tersebut, dimasukkan ke dalam data *grid* pada pgAdmin III mengantikan tingkat kepadatan sebelumnya. Penyatuan kembali dengan data *grid* awal bertujuan untuk mendapatkan kembali posisi lokasi masing-masing *grid* tersebut.

Setelah dimasukkan kembali ke dalam *grid*, tingkat kepadatan tersebut dikelompokan kembali ke dalam masing-masing kelas yang menghasilkan tabel kelas kerawanan rendah, kelas kerawanan sedang, dan kelas kerawanan tinggi. Pengelompokan tersebut bertujuan untuk mempermudah pewarnaan dalam memvisualisasikan ke dalam peta.

Untuk membangun peta kerawannya, ketiga data tersebut dimasukkan ke dalam QGIS. Setelah dimasukkan, masing-masing diberi warna yang berbeda, kelas rendah warna kuning, kelas sedang warna jingga, dan kelas tinggi warna merah. Peta Kabupaten Kubu Raya juga dimasukkan ke dalam peta kerawanan dengan warna hijau muda. Peta Kubu Raya tersebut bertujuan untuk memberikan informasi batas wilayah Kubu Raya. Gambar 8 merupakan peta tingkat kerawanan kebakaran hutan dan lahan.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Gambar 8 Peta tingkat kerawanan kebakaran hutan dan lahan Kabupaten Kubu Raya

SIMPULAN DAN SARAN

Simpulan

Penelitian ini berhasil mengklasifikasikan tingkat kepadatan titik panas dengan menggunakan tujuh attribut. Ketujuh atribut tersebut adalah jarak dengan jalan raya, jarak dengan hutan lahan kering, jarak dengan pusat pemerintahan, jarak dengan pemukiman warga, jarak dengan perkebunan, jarak dengan aliran sungai, dan jarak dengan lahan gambut.

Penggunaan teknik *oversampling* acak berhasil menghasilkan nilai akurasi, *precision* dan *recall* setiap kelas meningkat. Berdasarkan hasil nilai evaluasi tersebut, dapat disimpulkan ketujuh atribut yang digunakan bisa menjadi *rules* dalam menentukan kelas tingkat kepadatan titik panas wilayah Kabupaten Kubu Raya.

Model klasifikasi tingkat kepadatan titik panas memiliki korelasi positif terhadap tingkat kerawanan kebakaran hutan. Oleh karena itu, tingkat kepadatan titik panas dapat digunakan untuk memprediksi tingkat kerawanan kebakaran hutan dan lahan.

Penggunaan metode *k-fold cross validation* dengan $K = 10$ dan algoritme pohon keputusan C5.0 berhasil menghasilkan model pengklasifikasian tingkat kerawanan kebakaran hutan dalam bentuk *rules* dengan akurasi 90%. Atribut jarak dengan perkebunan menjadi atribut yang paling berpengaruh dan menjadi *root* pada model pohon keputusannya.

- Hak Cipta Dilindungi Undang-Undang
© Hak cipta milik IPB (Institut Pertanian Bogor)
Bogor Agricultural University
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Saran

Pengambilan data yang dilakukan belum mempertimbangkan beberapa hal seperti jarak sensor satelit dan faktor-faktor lainnya. Untuk itu, perlu adanya pengujian pada beberapa hal tersebut pada penelitian selanjutnya.

Prediksi kerawanan juga belum diujikan ke kondisi aktualnya, benar terjadi atau tidak kebakaran pada wilayah tersebut. Untuk itu, masih diperlukan perbaikan model agar mendapatkan hasil yang lebih baik.

DAFTAR PUSTAKA

- Chawla VN. 2003. C4.5 and Imbalanced Data sets Investigating the effect of sampling method, probabilistic estimate, and decision tree structure [Internet]; 2003 Agu 21; Toronto, Canada. Toronto (CA). [diunduh 2017 Mar 27]. Tersedia pada: <http://citeseerx.ist.psu.edu/messages/downloadsexceeded.html>
- Hadi M. 2006. *Pemodelan spasial kerawanan kebakaran di lahan gambut: studi kasus Kabupaten Bengkalis, Provinsi Riau* [tesis]. Bogor(ID): Institut Pertanian Bogor.
- Hajeri, Yurisithae E, dan Dolorosa E. 2015. *Analisis Penentuan Sektor Unggulan Perconomian di Kabupaten Kubu Raya*. Madura (ID). Fakultas Ekonomi dan Bisnis Universitas Trunojoyo.
- Han J, Kamber M, Pei J. 2012. *Data Mining Concepts and Techniques Third Edition*. San Francisco (US): Morgan Kaufmann Publisher.
- Jaya INS, Purnama ES, Arianti I, Boonyanuphap J. 2007. *Forest Fire Risk Assessment Model and post-Fire Evaluation using Remote Sensing and GIS: A Case Study in Riau, West Kalimantan and East Kalimantan Provinces, Indonesia*. Bogor(ID). Fakultas Kehutanan Institut Pertanian Bogor.
- Mallynur RU. 2011. *Penilaian risiko kebakaran lahan gambut di sumatera* [skripsi]. Bogor(ID): Institut Pertanian Bogor.
- Margono. 2004. *Metodologi Penelitian Pendidikan*. Jakarta (ID): Rineka Cipta.
- Munawaroh H, Khusnul B, dan Kustyaningsih Y. 2013. *Perbandingan Algoritme ID3 dan C5.0 Dalam Identifikasi Penjurusan Siswa SMA*. Madura (ID). Teknik Informatika Universitas Trunojoyo.
- Patil N, Lathi R, Chitre V. 2012. Customer card classification based on C5.0 and CART algorithms. *International Journal of Engineering Research and Applications*; 2012 July-August. 2 (4):164-167.
- Rulequest. 2012. *C5.0: An Informal Tutorial*. [Internet]. [diunduh 2017 Januari 12]. Tersedia pada: <https://www.rulequest.com/see5-unix.html>.
- Samsuri. 2008. *Model spasial tingkat kerawanan kebakaran hutan dan lahan studi kasus di wilayah provinsi Kalimantan Tengah* [tesis]. Bogor(ID): Institut Pertanian Bogor.



Zhao Y, Cao J, Zhang C, Zhang S. 2011. Enhancing grid-density based clustering for high dimensional data. *The Journal of System and Software*.doi:10.1016/j.jss.2011.02.047 hlm:1524-1539.

© Hak cipta milik IPB (Institut Pertanian Bogor)

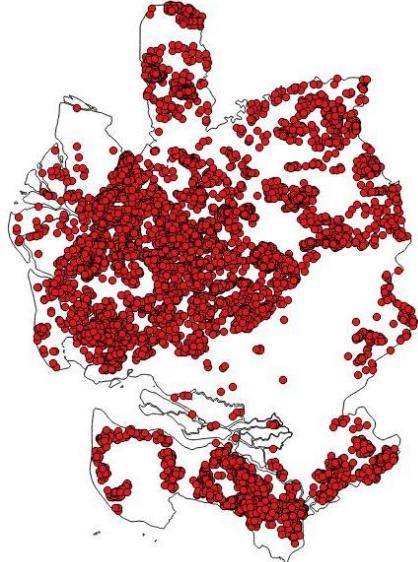
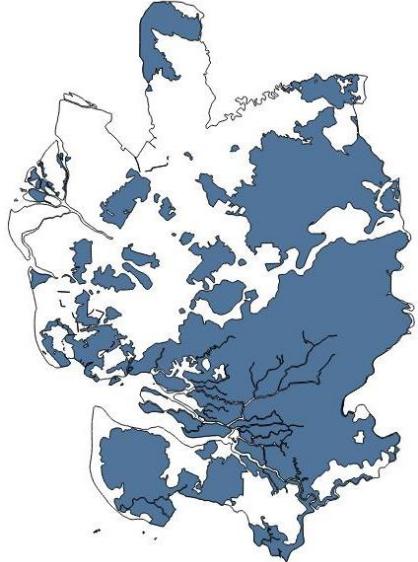
Bogor Agricultural U

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

LAMPIRAN

Lampiran 1 Data spasial yang digunakan

Nama Data Spasial	Sumber Data Spasial	Tipe Data Spasial	Data Spasial
Titik Panas	NASA	Point	
Hutan Kering	BIG	Polygon	

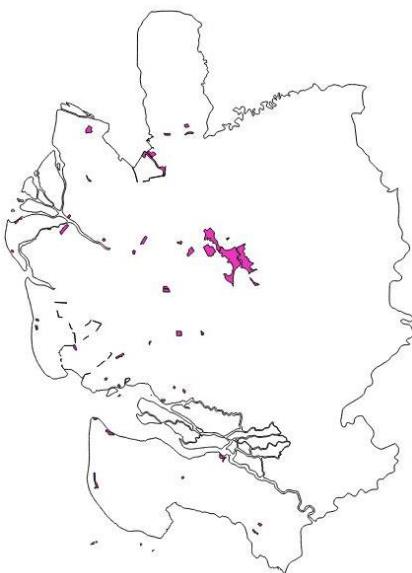
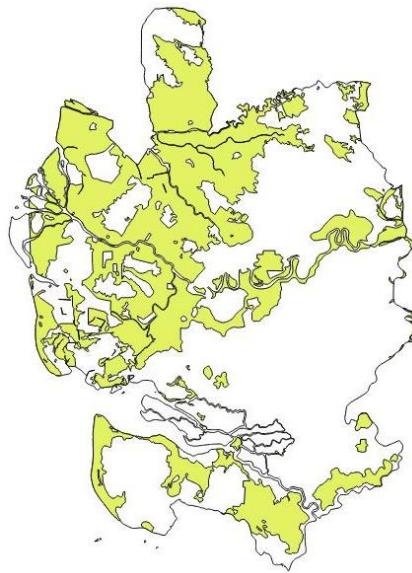
© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural U

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 1 Lanjutan

Nama Data Spasial	Sumber Data Spasial	Tipe Data Spasial	Data Spasial
Pemukiman Penduduk	BIG	Polygon	
Pelkebunan	BIG	Polygon	

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural U

Hak Cipta Dilindungi Undang-Undang

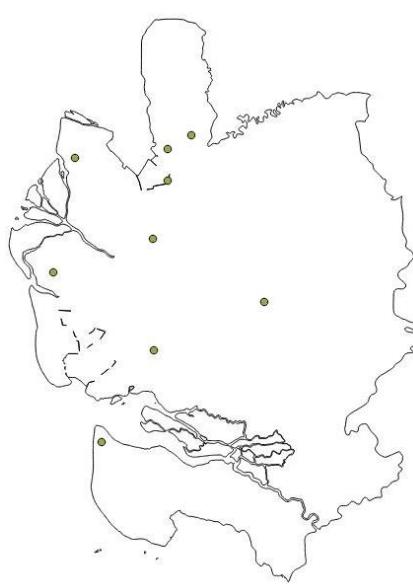
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 1 Lanjutan

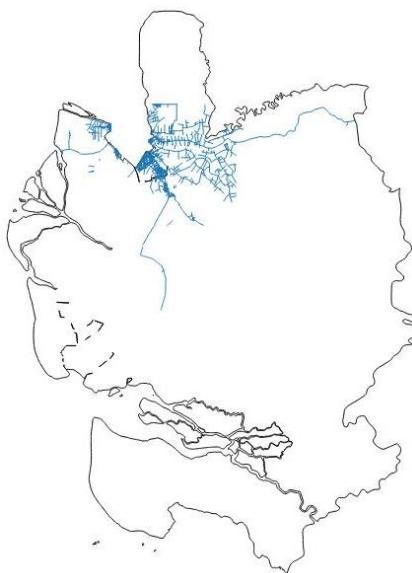
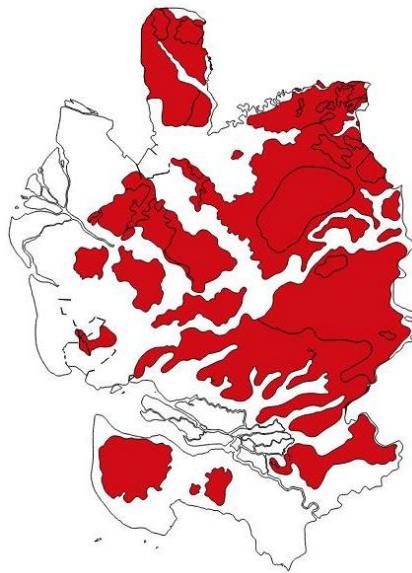
Nama Data Spasial	Sumber Data Spasial	Tipe Data Spasial	Data Spasial
Sungai	BIG	Polyline	
Pusat Pemerintahan	BIG	Point	

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural U

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 1 Lanjutan

Nama Data Spasial	Sumber Data Spasial	Tipe Data Spasial	Data Spasial
Jalan Raya	Geofabrik	Line	
Gambut	Wetlands International Indonesia	Polygon	

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 2 Kueri PostgreSQL untuk mendapatkan data jarak terpendek masing-masing atribut

```
create table Data as
select row_number() over (order by grid), kelas.grid, desa.kecamatan,
desa.desa,
(select st_distance(a.centroid, b.geom)*111.322 as jarak from kelas a,
jalanrayakuburaya b
where centroid = kelas.centroid
order by jarak limit 1) as JrkJalanRaya,
(select st_distance(a.centroid,b.geom)*111.322 as jarak from kelas a,
hutanlahankeringpolygon b
where centroid = kelas.centroid
order by jarak limit 1) as JrkHutanLahanKering,
(select st_distance(a.centroid,b.geom)*111.322 as jarak from kelas a,
pemerintahan b
where centroid = kelas.centroid
order by jarak limit 1) as JrkPemerintahan,
(select st_distance(a.centroid,b.geom)*111.322 as jarak from kelas a,
pemukiman b
where centroid = kelas.centroid
order by jarak limit 1) as JrkPemukiman,
(select st_distance(a.centroid,b.geom)*111.322 as jarak from kelas a,
pemukimanpenduduk b
where centroid = kelas.centroid
order by jarak limit 1) as JrkPerkebunan,
(select st_distance(a.centroid,b.geom)*111.322 as jarak from kelas a,
sungaipolygon b
where centroid = kelas.centroid
order by jarak limit 1) as JrkSungai,
(select st_distance(a.centroid,b.geom)*111.322 as jarak from kelas a,
gambut b
where centroid = kelas.centroid
order by jarak limit 1) as JrkGambut,
kelas.density
from desa cross join kelas where st_contains(desa.geom, kelas.centroid)
```

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Hak Cipta Dilindungi Undang-Undang



Lampiran 3 Tabel selang jarak masing-masing atribut

Jarak	Sangat	X > 13.086	X > 22.890	X > 10.668	X > 44.451	X > 56.915	X > 7.849	X > 53.286
Jauh	8.724 < X <= 15.260	7.630 < X <= 15.260	7.112 < X <= 11.113	11.113 < X <= 14.229	14.229 < X <= 22.226	22.226 < X <= 44.451	44.451 < X <= 56.915	56.915 < X <= 35.524 < X
Sedang	4.362 < X <= 8.724	7.630 < X <= 15.260	7.112 < X <= 11.113	11.113 < X <= 14.229	14.229 < X <= 22.226	22.226 < X <= 35.523	35.523 < X <= 7.849	>= 7.849 < X <= 35.524 < X
Dekat	X <= 4.362	X <= 7.630	X <= 3.556	X <= 11.113	X <= 2.616	X <= 2.616 < X <= 17.762	17.762 < X <= 35.524 < X	>= 35.524 < X
Berada	X = 0	X = 0	X = 0	X = 0	X = 0	X = 0	X = 0	X = 0
Ketinggian	-	-	-	-	-	-	-	-

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Lampiran 4 Baris kode R untuk memperoleh model beserta informasi lainnya

```
library("C50", lib.loc = "~/R/win-library/3.3")
library("plyr", lib.loc = "~/R/win-library/3.3")
library("caret", lib.loc = "~/R/win-library/3.3")
Data <- Data[sample(nrow(Data)),]
Tinggi <- Data[Data$density == 'Tinggi', ]
Sedang <- Data[Data$density == 'Sedang', ]
Rendah <- Data[Data$density == 'Rendah', ]
dataTinggi <- Data[sample(nrow(Tinggi)),]
foldsTinggi <- cut(seq(1, nrow(dataTinggi)), breaks = 10, labels = FALSE)
dataSedang <- Data[sample(nrow(Sedang)),]
foldsSedang <- cut(seq(1, nrow(dataSedang)), breaks = 10, labels = FALSE)
dataRendah <- Data[sample(nrow(Rendah)),]
foldsRendah <- cut(seq(1, nrow(dataRendah)), breaks = 10, labels = FALSE)
vektorAkurasi <- vector(length = 10)
for(i in 1:10){
  indexTinggi <- which(foldsTinggi == i, arr.ind = TRUE)
  indexSedang <- which(foldsSedang == i, arr.ind = TRUE)
  indexRendah <- which(foldsRendah == i, arr.ind = TRUE)
  testDataTinggi <- Tinggi[indexTinggi, ]
  trainDataTinggi <- Tinggi[-indexTinggi, ]
  testDataSedang <- Sedang[indexSedang, ]
  trainDataSedang <- Sedang[-indexSedang, ]
  testDataRendah <- Rendah[indexRendah, ]
  trainDataRendah <- Rendah[-indexRendah, ]
  testData <-
  join_all(list(testDataRendah, testDataSedang, testDataTinggi), type =
  'full')
  trainData <-
  join_all(list(trainDataRendah, trainDataSedang, trainDataTinggi), type =
  'full')
  assign(paste0("dataTes", i), testData)
  assign(paste0("dataTrain", i), trainData)
  modelnya <- C5.0(trainData[-8], trainData$density)
  assign(paste0("Model", i), modelnya)
  prediksinya <- (predict(modelnya, testData))
  temp <- table(prediksinya, testData$density)
  akurasi <- ((temp[1,1]+temp[2,2]+temp[3,3])/sum(temp)) * 100
  vektorAkurasi[i] <- akurasi
  assign(paste0("Hasil", i), confusionMatrix(prediksinya,
  testData$density))
}
rataRataAkurasi <- mean(vektorAkurasi)
a <- which.max(vektorAkurasi)
modelTerbaik <- paste0("Model", a)
rataRataAkurasi
modelTerbaik
```

Lampiran 5 Contoh perhitungan *information gain*

jarak jalan raya	jarak hutan lahan kering	jarak pemerintahan	jarak pemukiman	jarak perkebunan	jarak sungai	jarak gambut	density
Dekat	Sedang	Dekat	Dekat	Berada	Dekat	Dekat	Rendah
Sedang	Berada	Sedang	Jauh	Dekat	Sangat Jauh	Berada	Rendah
Sedang	Berada	Dekat	Sangat Jauh	Jauh	Dekat	Dekat	Sedang
Jauh	Sedang	Sangat Jauh	Dekat	Berada	Sedang	Sedang	Tinggi
Dekat	Dekat	Sedang	Berada	Dekat	Dekat	Berada	Sedang

Perhitungan *information gain* kelas

$$P(\text{Rendah}) = 2$$

$$P(\text{Sedang}) = 2$$

$$P(\text{Tinggi}) = 1$$

$$I_{\text{Kelas}} = -(P(\text{Rendah}) \log_2(\text{Rendah})) - (P(\text{Sedang}) \log_2(\text{Sedang})) - (P(\text{Tinggi}) \log_2(\text{Tinggi}))$$

$$I_{\text{Kelas}} = -\frac{2}{5} \log_2 \left(\frac{2}{5} \right) - \frac{2}{5} \log_2 \left(\frac{2}{5} \right) - \frac{1}{5} \log_2 \left(\frac{1}{5} \right) = 1,525$$

Perhitungan *information gain* atribut jalan raya

$$P(\text{Jalan Raya kategori Dekat}) = 2/5$$

Jumlah kelas rendah pada tuple atribut jalan raya kategori dekat = 1

Jumlah kelas sedang pada tuple atribut jalan raya kategori dekat = 1

Jumlah kelas tinggi pada tuple atribut jalan raya kategori dekat = 0

$$P(\text{Jalan Raya kategori Sedang}) = 2/5$$

- Jumlah kelas rendah pada tuple atribut jalan raya kategori Sedang = 1

- Jumlah kelas sedang pada tuple atribut jalan raya kategori Sedang = 1

- Jumlah kelas tinggi pada tuple atribut jalan raya kategori Sedang = 0

$$P(\text{Jalan Raya kategori Jauh}) = 2/5$$

- Jumlah kelas rendah pada tuple atribut jalan raya kategori Sedang = 0

- Jumlah kelas sedang pada tuple atribut jalan raya kategori Sedang = 0

- Jumlah kelas tinggi pada tuple atribut jalan raya kategori Sedang = 1

Menghitung *information gain* pada atribut jalan raya kategori dekat

$$I_{\text{Dekat}} = -\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{0}{2} \log_2 \left(\frac{0}{2} \right) = 1$$

$$I_{\text{Sedang}} = -\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{0}{2} \log_2 \left(\frac{0}{2} \right) = 1$$

$$I_{\text{Tinggi}} = -\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{0}{2} \log_2 \left(\frac{0}{2} \right) - \frac{0}{2} \log_2 \left(\frac{0}{2} \right) = 0.5$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Lampiran 5 Lanjutan

$$Gain_{(Dekat)} = 1,525 - \frac{2}{5}(1) = 1,125$$

$$Gain_{(Sedang)} = 1,525 - \frac{2}{15}(1) = 1,125$$

$$Gain_{(Tinggi)} = 1,525 - \frac{2}{5}(0.5) = 1,325$$

Karena *information gain* tertinggi dimiliki oleh kategori Tinggi, maka yang menjadi parent adalah atribut jalan raya kategori Tinggi



Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural University

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 6 *Confusion matrix* data uji sebelum di-oversampling

		Aktual					Aktual		
		Rendah	Sedang	Tinggi	Prediksi	Rendah	Sedang	Tinggi	Prediksi
Fold 1	Prediksi	Rendah	244	15	2	Rendah	244	14	2
Rendah	Sedang	0	0	0	Sedang	0	0	0	
Tinggi	Rendah	0	0	0	Tinggi	0	0	0	
Fold 2	Prediksi	Rendah	244	15	2	Fold 7	Prediksi	Rendah	244
Rendah	Sedang	0	0	0	Rendah	15	2	Prediksi	
Tinggi	Rendah	0	0	0	Sedang	0	0	Tinggi	
Fold 3	Prediksi	Rendah	244	15	2	Fold 8	Prediksi	Rendah	244
Rendah	Sedang	0	0	0	Rendah	15	2	Prediksi	
Tinggi	Rendah	0	0	0	Sedang	0	0	Tinggi	
Fold 4	Prediksi	Rendah	244	15	2	Fold 9	Prediksi	Rendah	244
Rendah	Sedang	0	0	0	Rendah	15	2	Prediksi	
Tinggi	Rendah	0	0	0	Sedang	0	0	Tinggi	
Fold 5	Prediksi	Rendah	244	15	2	Fold 10	Prediksi	Rendah	244
Rendah	Sedang	0	0	0	Rendah	15	2	Prediksi	
Tinggi	Rendah	0	0	0	Sedang	0	0	Tinggi	

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 7 Baris kode R untuk *k-fold cross validation*

```

# Mengacak data
Data <- Data[sample(nrow(Data)),]

#Memilah-milah data menjadi tiga tabel sesuai dengan datanya masing-masing
Tinggi <- Data[Data$density == 'Tinggi', ]
Sedang <- Data[Data$density == 'Sedang', ]
Rendah <- Data[Data$density == 'Rendah', ]

#Mengacak kembali data yang sudah dipilah pada masing-masing kelas
dataTinggi <- Data[sample(nrow(Tinggi)),]
dataSedang <- Data[sample(nrow(Sedang)),]
dataRendah <- Data[sample(nrow(Rendah)),]

#Pengindexan pada setiap baris data dari 1 sampai 10 (menyesuaikan jumlah
#fold) secara bergantian
foldsTinggi <- cut(seq(1,nrow(dataTinggi)), breaks = 10, labels = FALSE)
foldsSedang <- cut(seq(1,nrow(dataSedang)), breaks = 10, labels = FALSE)
foldsRendah <- cut(seq(1,nrow(dataRendah)), breaks = 10, labels = FALSE)

#Proses penggunaan 10 Fold pada perulangan sebanyak 10 kali
for(i in 1:10){

  #Pemilihan index data sesuai index perulangan ke - i
  indexTinggi <- which(foldsTinggi==i, arr.ind = TRUE)
  indexSedang <- which(foldsSedang==i, arr.ind = TRUE)
  indexRendah <- which(foldsRendah==i, arr.ind = TRUE)

  #Data yang terpilih indexnya akan dijadikan sub data uji
  testDataTinggi <- Tinggi[indexTinggi, ]
  testDataSedang <- Sedang[indexSedang, ]
  testDataRendah <- Rendah[indexRendah, ]

  #Indexnya tidak terpilih dijadikan sebagai sub data latih
  trainDataTinggi <- Tinggi[-indexTinggi, ]
  trainDataSedang <- Sedang[-indexSedang, ]
  trainDataRendah <- Rendah[-indexRendah, ]

  #Penyatuan kembali masing-masing data menjadi data uji dan data latih
  testData <- jo
  n_all(list(testDataRendah,testDataSedang,testDataTinggi), type =
  'full'
  trainData <-
  join_all(list(trainDataRendah,trainDataSedang,trainDataTinggi), type =
  'full')
}

```

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Hak Cipta Dilindungi Undang-Undang

Lampiran 8 Pohon keputusan model 10 (model terbaik)

```

jrkperekunan in {Jauh,Sedang}:
....jrkpemukiman in {Berada,Jauh,Sangat Jauh,Sedang}: Sedang (441)
....jrkpemukiman = Dekat:
....jrksungai in {Berada,Jauh,Sangat Jauh,Sedang}: Sedang (84)
....jrksungai = Dekat:
....jrkjalanraya in {Jauh,Sangat Jauh,Sedang}: Sedang (19)
....jrkjalanraya = Dekat:
....jrkutanlahankering in {Berada,Dekat}: Rendah (6)
....jrkutanlahankering in {Jauh,Sangat Jauh,Sedang}: Sedang (3)
jrkperekunan in {Berada,Dekat}:
....jrkpemukiman = Sangat Jauh:
....jrkutanlahankering in {Jauh,Sangat Jauh}: Sedang (35)
....jrkutanlahankering in {Berada,Dekat,Sedang}:
....jrksungai in {Berada,Sangat Jauh}: Sedang (18)
....jrksungai in {Dekat,Jauh,Sedang}:
....jrkpemerintahan = Sedang: Sedang (8)
....jrkpemerintahan in {Dekat,Jauh,Sangat Jauh}:
....jrkgambut in {Berada,Dekat,Sangat Jauh,
....Sedang}: Tinggi (625/31)
....jrkgambut = Jauh: Sedang (6)
jrkpemukiman in {Berada,Jauh,Sedang}:
....jrkjalanraya in {Jauh,Sangat Jauh}: Sedang (157)
....jrkjalanraya in {Dekat,Sedang}:
....jrkgambut in {Jauh,Sangat Jauh,Sedang}: Sedang (78/3)
....jrkgambut in {Berada,Dekat}:
....jrkpemerintahan in {Jauh,Sangat Jauh}: Sedang (49/5)
....jrkpemerintahan in {Dekat,Sedang}:
....jrksungai = Berada: Sedang (12/1)
....jrksungai in {Dekat,Jauh,Sangat Jauh,Sedang}:
....jrkutanlahankering = Sangat Jauh: Sedang (7)
....jrkutanlahankering in {Jauh,Sedang}:
....jrkgambut = Berada: Sedang (16/5)
....jrkgambut = Dekat: Rendah (7/2)
....jrkutanlahankering in {Berada,Dekat}:
....jrkpemukiman in {Jauh,Sedang}: Rendah (76/14)
....jrkpemukiman = Berada:
....jrkutanlahankering = Berada: Sedang (3)
....jrkutanlahankering = Dekat: Rendah (12/5)

jrkpemukiman = Dekat:
....jrkpemerintahan in {Dekat,Sedang}:
....jrksungai = Sangat Jauh: Sedang (7)
....jrksungai in {Berada,Dekat}:
....jrkjalanraya = Sangat Jauh:
....jrksungai = Berada: Sedang (2)
....jrksungai = Dekat: Tinggi (17)
....jrkjalanraya = Jauh:
....jrkpemerintahan = Sedang: Rendah (25/2)
....jrkpemerintahan = Dekat:
....jrkutanlahankering in {Berada,Dekat,
....Sedang}: Tinggi (19/1)
....jrkutanlahankering in {Jauh,
....Sangat Jauh}: Sedang (3)
....jrkjalanraya in {Dekat,Sedang}:
....jrkjalanraya = Dekat:
....jrkgambut in {Berada,Dekat,Jauh,
....Sangat Jauh}: Rendah (829/87)
....jrkgambut = Sedang:

```

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik iPB (Institut Pertanian Bogor)

Bogor Agricultural

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 8 Lanjutan

```

:   :   ....jrkperkebunan = Berada: Rendah (19/2)
:   :   jrkperkebunan = Dekat:
:   :       ....jrkpemerintahan = Dekat: Tinggi (5/2)
:   :       jrkpemerintahan = Sedang: Sedang (2)
jrkjalanraya = Sedang:
....jrkgambut in {Dekat,Jauh,Sangat Jauh,
:           Sedang}: Rendah (148/13)
jrkgambut = Berada:
....jrkpemerintahan = Sedang: Rendah (12/4)
jrkpemerintahan = Dekat: [S1]

jrksungai = Jauh:
....jrkpemerintahan = Sedang: Sedang (4)
jrkpemerintahan = Dekat:
....jrkjalanraya in {Dekat,Sangat Jauh,
:           Sedang}: Tinggi (50/2)
jrkjalanraya = Jauh:
....jrhutanlahankering in {Jauh,
:           Sangat Jauh}: Rendah (0)
jrhutanlahankering = Sedang: Tinggi (5)
jrhutanlahankering in {Berada,Dekat}:
....jrkgambut = Berada: Rendah (17/2)
jrkgambut in {Jauh,Sangat Jauh,
:           Sedang}: Tinggi (3)
jrkgambut = Dekat:
....jrhutanlahankering = Berada: Tinggi (2)
jrhutanlahankering = Dekat: Rendah (9/2)

jrksungai = Sedang:
....jrkgambut = Sangat Jauh: Rendah (0)
jrkgambut in {Jauh,Sedang}:
....jrkpemerintahan = Dekat: Tinggi (22/2)
jrkpemerintahan = Sedang: Sedang (4)
jrkgambut in {Berada,Dekat}:
....jrkpemerintahan = Sedang:
....jrhutanlahankering in {Berada,Dekat,
:           Sangat Jauh}: Rendah (40/3)
:   jrhutanlahankering in {Jauh,Sedang}: Sedang (2)
jrkpemerintahan = Dekat:
....jrkjalanraya in {Jauh,Sangat Jauh}: Tinggi (21/3)
jrkjalanraya = Sedang:
....jrhutanlahankering in {Berada,Dekat,
:           Sangat Jauh}: Rendah (21/8)
:   jrhutanlahankering = Jauh: Sedang (1)
:   jrhutanlahankering = Sedang: Tinggi (4)
jrkjalanraya = Dekat:
....jrkgambut = Berada: Rendah (68/13)
jrkgambut = Dekat:
....jrkperkebunan = Berada: Rendah (21/8)
jrkperkebunan = Dekat: Tinggi (4/1)

jrkpemerintahan in {Jauh,Sangat Jauh}:
jrhutanlahankering in {Jauh,Sangat Jauh}: Sedang (29)
jrhutanlahankering in {Berada,Dekat,Sedang}:
....jrksungai in {Berada,Sangat Jauh}: Sedang (11)
jrksungai in {Dekat,Jauh,Sedang}:
....jrkjalanraya in {Dekat,Sangat Jauh,Sedang}:
....jrkgambut in {Berada,Dekat,Sangat Jauh,
:           Sedang}: Tinggi (311/21)
:   jrkgambut = Jauh: Sedang (2)

```

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural U

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 8 Lanjutan

jrkjalanraya = Jauh:
....jrksungai in {Jauh,Sedang}: Tinggi (74/7)
jrksungai = Dekat:
....jrkpemerintahan = Sangat Jauh: Tinggi (19/2)
jrkpemerintahan = Jauh:
....jrkgambut in {Jauh,
: Sangat Jauh}: Rendah (0)
jrkgambut = Sedang: Tinggi (8/1)
jrkgambut in {Berada,Dekat}: [S2]

Subtree [S1]

jrkutanlahankering = Berada: Rendah (6/2)
jrkutanlahankering in {Dekat,Jauh,Sangat Jauh}: Sedang (12/4)
jrkutanlahankering = Sedang: Tinggi (2)

Subtree [S2]

jrkutanlahankering = Berada: Rendah (17/4)
jrkutanlahankering = Sedang: Tinggi (4)
jrkutanlahankering = Dekat:
....jrkperkebunan = Berada: Rendah (7/3)
....jrkperkebunan = Dekat: Tinggi (2)

Attribute usage:

100.00%	jrkperkebunan
84.56%	jrkpemukiman
76.12%	jrksungai
75.84%	jrkpemerintahan
56.40%	jrkjalanraya
46.95%	jrkutanlahankering
45.23%	jrkgambut

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 9 *Confusion matrix* data uji setelah di-oversampling

		Aktual			Aktual		
		Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi
Prediksi	Rendah	129	15	8	129	18	6
	Sedang	2	105	0	2	107	2
	Tinggi	1	12	124	1	7	124
Fold 1		Aktual			Aktual		
Prediksi	Rendah	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi
	Rendah	131	20	14	128	19	11
	Sedang	1	106	0	2	107	1
Prediksi	Tinggi	0	6	118	1	5	119
Fold 2		Aktual			Aktual		
Prediksi	Rendah	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi
	Rendah	127	17	8	121	13	6
	Sedang	3	110	0	4	107	1
Prediksi	Tinggi	1	4	123	7	12	125
Fold 3		Aktual			Aktual		
Prediksi	Rendah	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi
	Rendah	127	12	8	121	17	9
	Sedang	5	110	2	7	109	0
Prediksi	Tinggi	0	10	122	3	5	122
Fold 4		Aktual			Aktual		
Prediksi	Rendah	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi
	Rendah	127	12	8	128	13	10
	Sedang	5	110	2	3	110	1
Prediksi	Tinggi	0	10	122	1	9	121
Fold 5		Aktual			Aktual		
Prediksi	Rendah	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi
	Rendah	124	14	10	128	13	10
	Sedang	7	105	0	3	110	1
Prediksi	Tinggi	0	12	121	1	9	121
Fold 6		Aktual			Aktual		
Prediksi	Rendah	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi
	Rendah	129	18	6	128	19	11
	Sedang	2	107	2	2	107	1
Prediksi	Tinggi	1	7	124	1	5	119
Fold 7		Aktual			Aktual		
Prediksi	Rendah	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi
	Rendah	128	19	11	121	13	6
	Sedang	2	107	1	4	107	1
Prediksi	Tinggi	1	5	119	7	12	125
Fold 8		Aktual			Aktual		
Prediksi	Rendah	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi
	Rendah	121	13	6	121	17	9
	Sedang	4	107	1	7	109	0
Prediksi	Tinggi	7	12	125	3	5	122
Fold 9		Aktual			Aktual		
Prediksi	Rendah	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi
	Rendah	121	17	9	121	17	9
	Sedang	7	109	0	7	109	0
Prediksi	Tinggi	3	5	122	3	5	122
Fold 10		Aktual			Aktual		
Prediksi	Rendah	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi
	Rendah	128	13	10	128	13	10
	Sedang	3	110	1	3	110	1
Prediksi	Tinggi	1	9	121	1	9	121



RIWAYAT HIDUP

Penulis dilahirkan di Taba Tembilang, Bengkulu pada tanggal 12 Februari 1996. Penulis merupakan anak ketiga dari tiga bersaudara. Penulis adalah anak kandung dari pasangan Suhardi dan Nelviarti.

Penulis menempuh pendidikan Sekolah Menengah Atas di SMAN 1 Arga Makmur dan lulus pada tahun 2013. Kemudian penulis melanjutkan jenjang pendidikan di Departemen Ilmu Komputer Fakultas Matematika dan IPA Institut Pertanian Bogor (IPB) dari tahun 2013 hingga tahun 2017. Selama menempuh jenjang perkuliahan, penulis berhasil meraih prestasi dalam kompetisi *data science* Gemastik 9 di Universitas Indonesia sebagai finalis pada tahun 2016. Penulis juga aktif sebagai staff Himpunan Mahasiswa Ilmu Komputer 2015, dan wakil ketua Himpunan Mahasiswa Ilmu Komputer 2016. Selain itu pada bulan Juli dan Agustus 2016 penulis melaksanakan kegiatan Praktik Kerja Lapang di *Division of Information Technology* Asosiasi Asuransi Jiwa Indonesia Jakarta Pusat.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.