

Pengunaan Teknik *Data Mining* dalam Pemodelan Resiko Terjadinya Kebakaran Hutan

Imas Sukaesih Sitanggang
Departemen Ilmu Komputer, FMIPA, Institut Pertanian Bogor, Indonesia
e-mail: imas.sitanggang@ipb.ac.id

ABSTRAK

Kebakaran hutan di Indonesia telah mengakibatkan banyak kerugian dalam berbagai aspek kehidupan. Pencegahan terjadinya kebakaran hutan sangat diperlukan untuk meminimalkan kerugian yang diakibatkan bencana tersebut. Upaya pencegahan dapat dilakukan melalui pemodelan untuk menilai resiko terjadinya kebakaran hutan. Pemodelan resiko kebakaran hutan telah banyak dilakukan menggunakan sistem informasi geografis dan penginderaan jarak jauh. Tulisan ini menyajikan tinjauan terhadap penggunaan teknik *data mining* dalam pemodelan resiko kebakaran hutan. Teknik yang digunakan adalah klasifikasi dengan menggunakan algoritme pohon keputusan C4.5, SimpleCart, NaïveBayes, regresi logistik, dan pohon keputusan ID3 spasial. Akurasi tertinggi untuk model yang dihasilkan mencapai 87.69%. Model resiko kebakaran dapat digunakan untuk memprediksi kemunculan titik api di lokasi baru berdasarkan karakteristik lokasi tersebut.

Kata kunci: data mining; kebakaran hutan; teknik klasifikasi

I. PENDAHULUAN

Pencegahan kebakaran hutan perlu dilakukan untuk meminimalkan kerugian akibat bencana tersebut. Salah satu upaya yang dilakukan dalam pencegahan kebakaran hutan adalah menentukan resiko terjadinya kebakaran hutan. Teknologi berbasis komputer meliputi sistem informasi geografis (SIG) dan penginderaan jarak jauh telah banyak digunakan dalam pemodelan resiko dan bahaya kebakaran hutan. Pemodelan bahaya kebakaran hutan di Kalimantan Timur menggunakan penginderaan jarak jauh yang diintegrasikan dengan SIG telah dilakukan dalam studi oleh Darmawan et.al. (2000) [1]. Boonyanuphap (2001) dalam [2] membangun model resiko kebakaran hutan untuk daerah Sasamba Kalimantan Indonesia. Metode yang digunakan adalah Complete Mapping Analysis (CMA) untuk menganalisis kerentanan kebakaran berdasarkan faktor lingkungan fisik yaitu rata-rata temperatur minimum harian, curah hujan, kelembaban relatif, *agro-*

climatic zone, kemiringan, serta faktor aktifitas manusia yaitu pusat desa, jalan, vegetasi dan penutupan lahan. Studi lain dilakukan oleh Hadi (2006) [3] untuk memodelkan resiko kebakaran di lahan gambut di Bengkalis Provinsi Riau berdasarkan distribusi hotspot sebagai indikator terjadinya kebakaran hutan, aspek lingkungan meliputi penutupan lahan, tipe vegetasi, kedalaman gambut, tingkat kehijauan vegetasi, dan aspek infrastruktur meliputi jarak kemunculan api dengan jalan. Pembobotan dan penentuan skor untuk setiap faktor tersebut dilakukan menggunakan metode Complete Mapping of Analysis (CMA).

Di samping teknologi SIG dan penginderaan jarak jauh, saat ini penerapan teknik-teknik *data mining* dalam pemodelan resiko kebakaran hutan mulai berkembang khususnya ketika pemodelan tersebut melibatkan banyak faktor dan data berukuran besar. *Data mining* merupakan proses ekstraksi informasi yang berguna dalam

data ukuran besar. Teknik-teknik klasifikasi, *clustering* dan aturan asosiasi dalam *data mining* diaplikasikan dalam data kebakaran hutan untuk mencari pola-pola menarik mengenai keterkaitan faktor menentu terjadinya kebakaran hutan yang berguna dalam pengambilan keputusan terkait pencegahan kebakaran hutan. Tulisan ini menyajikan tinjauan untuk beberapa studi dalam penerapan teknik-teknik *data mining* dalam data kebakaran hutan spasial. Tulisan ini diorganisasikan sebagai berikut. Subbab selanjutnya menjelaskan beberapa studi penerapan teknik *data mining* dalam pemodelan resiko kebakaran hutan yang telah dilakukan dalam beberapa studi. Salah satu metode klasifikasi yaitu pohon keputusan dan aplikasinya pada data kebakaran hutan dijelaskan dalam Subbab 3. Ringkasan tulisan diberikan pada subbab terakhir.

II. PENGGUNAAN TEKNIK DATA MINING DALAM DATA KEBAKARAN HUTAN

Data kebakaran hutan umumnya berupa data spasial. Data spasial memiliki karakteristik spasial dan non spasial. Atribut spasial digunakan untuk mendefinisikan lokasi objek. Sedangkan atribut non spasial menjelaskan karakteristik objek seperti nama dan jumlah populasi untuk area tertentu. Teknik-teknik *data mining* telah digunakan dalam mencari pola menarik dari data spasial berukuran besar. Proses tersebut dinyatakan sebagai *spatial data mining* dimana algoritme *data mining* spasial dan non-spasial diimplementasikan pada data spasial. Penerapan teknik *data mining* dalam data spasial lebih kompleks dari penggunaan teknik tersebut pada data non spasial, karena algoritme-algoritme *data mining* spasial harus memperhitungkan objek tetangga dari objek studi untuk mendapatkan pola yang menarik. Dalam *data mining* spasial, atribut-atribut dari tetangga sebuah objek dapat memberikan

pengaruh yang signifikan kepada objek studi.

Teknik-teknik *data mining* telah banyak diterapkan dalam menganalisis data spasial untuk kebakaran hutan. Prasad dan Ramakrishna [4] menggunakan algoritme *clustering* yaitu K-means dan logika fuzzy dalam penentuan lokasi yang memiliki resiko terjadinya kebakaran hutan. Studi tersebut mengajukan sebuah sistem untuk mengidentifikasi kebakaran hutan dari citra satelit [4]. Stojanova et. al. [5] telah membangun model prediktif berdasarkan data dari satelit meteorologi ALADIN dan MODIS. Teknik *data mining* yang digunakan adalah regresi logistik, pohon keputusan, *random forests*, *bagging* dan *boosting*, untuk memperoleh model prediktif untuk kemunculan kebakaran.

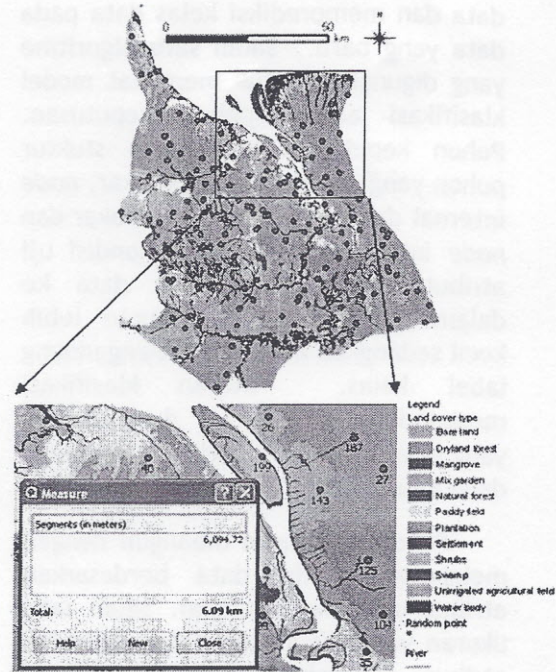
Dalam studi oleh Cortez dan Morais [6], teknik lainnya dalam *data mining* yaitu Support Vector Machines dan *Random Forests* diterapkan untuk memprediksi area yang terbakar karena kebakaran hutan berdasarkan faktor meteorologi di area timur laut di Portugal, yaitu temperature, kelembaban relative, curah hujan dan kecepatan angin. Angayarkkani dan Radhakrishnan (2009) [7] menentukan area yang rawan terjadinya kebakaran hutan dengan menggunakan teknik *data mining* spasial dan logika fuzzy. Teknik *data mining* lain yang digunakan dalam analisis data kebakaran hutan adalah algoritme Apriori untuk mencari aturan asosiasi [8]. Algoritme tersebut digunakan untuk menentukan peluang dan intensitas kebakaran hutan berdasarkan data cuaca.

III. PENGGUNAAN ALGORITME POHON KEPUTUSAN UNTUK PEMODELAN RESIKO KEBAKARAN HUTAN

Teknik klasifikasi dalam *data mining* digunakan untuk membuat model yang menjelaskan kelas-kelas

4. IF penutupan lahan = Unirrigated_agricultural_field AND jarak ke sungai terdekat > 353.66 meter THEN Kemunculan Titik Api = T (40.0/19.0)
5. IF penutupan lahan = Dryland_forest AND jarak ke pusat kota terdekat <= 14807.65 meter THEN Kemunculan Titik Api = F (77.0/27.0)

Bilangan dalam tanda kurung pada setiap aturan di atas menyatakan banyaknya objek dalam *node* daun yang bersesuaian dengan aturan, dimana banyaknya objek yang salah diklasifikasikan oleh model diberikan setelah tanda /. Dari keseluruhan objek dalam dataset, terdapat 470 *tuple* yang diklasifikasikan benar oleh pohon keputusan. Selanjutnya pohon keputusan yang dihasilkan dapat digunakan untuk memprediksikan kemunculan titik api pada lokasi baru. Sebagai contoh, pada Gambar 1, sebuah titik no 187 berada pada lokasi perkebunan (*plantation*) dan lokasi tersebut berada 6,09 km dari sungai terdekat. Berdasarkan aturan no 2 dapat dinyatakan bahwa lokasi tersebut diklasifikasikan ke dalam kelas positif dimana di lokasi tersebut dimungkinkan terjadinya titik api.



Gambar 1. Jarak lokasi titik 187 ke sungai terdekat.

Sitanggang et al [11] telah menerapkan beberapa algoritme klasifikasi yang tersedia dalam *data mining* toolkit Weka 3.6.2 pada data kebakaran hutan Provinsi Riau. Algoritme tersebut adalah pohon keputusan C4.5, SimpleCart dan NaïveBayes. Data spasial yang digunakan adalah penutupan lahan, jalan, sungai, pusat kota, Hak Pengusahaan Hutan (HPH), Hutan Tanaman Industri (HTI), kedalaman dan jenis lahan gambut, untuk mengklasifikasikan 2693 objek target. Target objek terdiri dari data dengan kelas positif (*true alarm data*) yang merupakan hotspot pada 2008 dan data dengan kelas negative (*false alarm data*) yang dibangkitkan secara acak dalam area yang berjarak sedikitnya 1 km dari *true alarm data*. Hasil menunjukkan bahwa model klasifikasi yang dihasilkan oleh algoritme pohon keputusan C4.5 memiliki akurasi paling tinggi yaitu 69.59 % jika dibandingkan dengan dua algoritme yang lain.

Berdasarkan pohon keputusan yang diperoleh menggunakan algoritme C4.5, titik api terjadi di area yang memiliki

data dan memprediksi kelas data pada data yang baru. Salah satu algoritme yang digunakan untuk membuat model klasifikasi adalah pohon keputusan. Pohon keputusan merupakan struktur pohon yang terdiri dari *node* akar, *node* internal dan *node* daun. *Node* akar dan *node* internal mengandung kondisi uji atribut untuk memisahkan data ke dalam partisi-partisi berukuran lebih kecil sedangkan *node* daun mengandung label kelas. Aturan klasifikasi mengandung atribut uji dan nilainya yang diturunkan dari *node* akar ke *node* daun dalam pohon keputusan.

Pohon keputusan dibangun dengan melakukan partisi data berdasarkan atribut uji yang dijumpai. Salah satu ukuran yang digunakan dalam memilih atribut pemisah data adalah *Information gain*. Misalkan *node* N menyatakan *tuple* dalam partisi D. Atribut yang memiliki nilai *information gain* tertinggi dipilih sebagai atribut pemisah untuk *node* N. Atribut terpilih meminimumkan informasi yang diperlukan untuk mengklasifikasikan *tuple* dalam partisi-partisi yang dihasilkan [9]. Informasi yang diperlukan untuk mengklasifikasikan sebuah *tuple* dalam D ditentukan sebagai berikut:

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

dimana p_i adalah peluang bahwa sebuah *tuple* dalam D memiliki label kelas C_i , dimana nilai peluang ini diestimasi dengan $|C_{i,D}|/|D|$. $Info(D)$ dikenal sebagai entropy dari D. Diasumikan bahwa kita ingin mempartisi *tuple-tuple* di D pada atribut A yang memiliki sebanyak v nilai yang berbeda, $\{a_1, a_2, \dots, a_v\}$. Partisi-partisi yang dihasilkan berkaitan dengan cabang-cabang dari *node* N. $Info_A(D)$ merupakan informasi yang diinginkan untuk mengklasifikasikan *tuple* dari D berdasarkan pemisahan data oleh A [9]. Berikut adalah formula $Info_A(D)$ [9]:

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

dimana $\frac{|D_j|}{|D|}$ merupakan bobot dari partisi ke j . *Information gain* dinyatakan sebagai selisih antara informasi awal yang diperlukan berdasarkan proporsi kelas dan informasi yang diperlukan untuk pemisahan data berdasarkan atribut A. Formula *Information gain* adalah sebagai berikut [9].

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

Selanjutnya atribut A dengan nilai *information gain* tertinggi dipilih sebagai atribut pemisah pada *node* N.

Algoritme pohon keputusan yang banyak digunakan diantaranya adalah ID3, C4.5 dan CART. Algoritma C4.5 telah digunakan dalam menentukan model klasifikasi untuk dataset kebakaran hutan di Kabupaten Rokan Hilir, Provinsi Riau [10]. Data spasial yang digunakan meliputi pusat kota, jaringan jalan, sungai dan penutupan lahan. Operasi spasial digunakan untuk menghubungkan data spasial tersebut dengan titik api. Dataset yang dihasilkan dari operasi spasial terdiri dari 744 *tuple* (374 kelas positif dan 370 kelas negatif). Model klasifikasi berupa pohon keputusan dengan 18 *node* daun dan 26 *node* internal dengan akurasi 63.17%. *Node* daun menyatakan kelas data yaitu kelas positif (T) dan kelas negatif (F). Berikut adalah beberapa aturan yang dihasilkan

1. IF penutupan lahan = Plantation AND jarak ke sungai terdekat \leq 4546.97 meter THEN Kemunculan Titik Api = F (187.0/76.0)
2. IF penutupan lahan = Plantation AND jarak ke sungai terdekat $>$ 4546.97 meter THEN Kemunculan Titik Api = T (125.0/30.0)
3. IF penutupan lahan = Shrubs THEN Kemunculan Titik Api = F (63.0/28.0)

karakteristik sebagai berikut: 1) bukan area HPH, 2) merupakan area perkebunan dan *dryland forest*, 3) tipe lahan gambut Hemists/Saprists dengan kedalaman di atas 400 cm, 4) area dengan jarak ke jalan terdekat mencapai 10 km, jarak ke sungai terdekat mencapai 3 km, 5) area tersebut berada diantara 5 km sampai dengan 20 km dari pusat kota.

Di samping algoritme pohon keputusan, algoritme lain yang dapat digunakan dalam membuat model klasifikasi adalah regresi logistik. Studi yang dilakukan oleh Sitanggang et al [12] menerapkan regresi logistik pada data kebakaran hutan untuk memprediksi kemunculan titik api berdasarkan karakteristik fisik, sosial-ekonomi dan cuaca untuk wilayah Rokan Hilir. Model regresi logistik yang dihasilkan terdiri dari variabel jarak ke pusat kota terdekat, jarak ke sungai terdekat, jarak ke jalan terdekat, tipe lahan gambut, curah hujan, temperatur dan kecepatan angin, untuk mengklasifikasikan objek ke dalam kelas positif (kemunculan titik api) atau kelas negatif (bukan kemunculan titik api).

Pengembangan algoritme pohon keputusan ID3 telah dilakukan oleh Sitanggang et al [13] sehingga algoritme tersebut dapat langsung digunakan dalam mengelolah data spasial yang berisi fitur diskret (titik, garis dan poligon). Dalam klasifikasi data spasial, dataset merupakan kumpulan dari layer dimana setiap fitur spasial dalam suatu layer memiliki tipe geometri yang sama (titik, garis, dan poligon). Terdapat dua jenis layer dalam dataset spasial yaitu layer penjelas dan layer target (layer referensi) yang menyimpan kelas target. Layer penjelas memiliki beberapa atribut, dimana salah satunya adalah atribut prediktor yang akan mengklasifikasikan *tuple* dalam dataset ke dalam kelas target. Relasi spasial diaplikasikan kepada layer-layer tersebut untuk membentuk sejumlah *tuple* dalam layer-layer baru. Relasi

spasial tersebut mencakup *topological* (seperti *overlap*, *intersection*) dan *metric* (jarak).

Seperti halnya algoritme ID3 yang menggunakan *information gain* dalam seleksi atribut, algoritme ID3 spasial menggunakan *spatialinformation gain* untuk memilih layer pemisah terbaik dari sejumlah layer dalam basis data spasial. Formula baru untuk *spatialinformation gain* diajukan dengan menggunakan ukuran-ukuran spasial untuk fitur titik, garis dan poligon.

Spatialinformation gain dihitung berdasarkan nilai entropy sebelum pemisahan dataset dan setelah pemisahan dataset oleh layer tertentu. Formula entropy dinyatakan sebagai berikut. Misalkan atribut target C dalam layer target S memiliki l kelas yang berbeda (yaitu c_1, c_2, \dots, c_l), entropy untuk S merepresentasikan informasi yang diperlukan untuk menentukan kelas dari *tuple* dalam dataset dan didefinisikan sebagai berikut [13]:

$$H(S) = - \sum_{i=1}^l \frac{SpatMes(S_{c_i})}{SpatMes(S)} \log_2 \frac{SpatMes(S_{c_i})}{SpatMes(S)} \quad (4)$$

$SpatMes(S)$ adalah ukuran spasial dari layer S yang dapat berupa area ataupun jarak yang dihitung dengan mengaplikasikan relasi spasial antar dua layer yang berbeda. Misalkan suatu atribut penjelas V dalam layer penjelas (non-target) L memiliki q nilai yang berbeda (yaitu v_1, v_2, \dots, v_q). Partisi objek dalam layer target S berdasarkan layer L akan menghasilkan sejumlah layer baru $L(v_i, S)$ untuk setiap nilai v_i dalam L. Dengan asumsi bahwa layer L menutup semua area dalam layer S, nilai *expected entropy* untuk pemisahan tersebut diberikan sebagai berikut:

$$H(S|L) = \sum_{j=1}^q \frac{SpatMes(L(v_j, S))}{SpatMes(S)} H(L(v_j, S)) \quad (5)$$

H(S|L) menyatakan jumlah informasi yang diperlukan (setelah pemisahan dataset). Selanjutnya nilai spatial *information gain* untuk layer L diberikan oleh formula berikut

$$\text{Gain}(L) = H(S) - H(S|L) \quad (6)$$

Nilai Gain(L) menunjukkan besarnya informasi yang dapat diperoleh oleh pencabangan pohon keputusan pada layer L. Layer L dengan nilai *information gain* tertinggi dipilih sebagai layer pemisah dari dataset pada suatu *node* N.

Algoritme ID3 spasial telah diaplikasikan ada data spasial kebakaran hutan yang berisi fitur-fitur titik, garis, dan poligon untuk karakteristik area di kabupaten Rokan Hilir, Provinsi Riau. Data spasial mencakup karakteristik fisik, social-ekonomi dan cuaca, serta kemunculan titik api sebagai objek target (Tabel 1).

Tabel 1. Layer dan Banyaknya Fitur dalam Dataset

Layer	Banyaknya fitur
Fisik	
Jarak ke sungai terdekat (dist_river)	744 points
Jarak ke jalan terdekat (dist_road)	744 points
Jarak ke pusat kota terdekat (dist_city)	744 points
Penutupan lahan (land_cover)	3107 polygons
Social-ekonomi	
Sumber pendapatan (income_source)	117 polygons
Kepadatan populasi (population)	117 polygons
Jumlah sekolah (school)	117 polygons
Cuaca	
Curah hujan dalam mm/day (precipitation)	7 polygons
Temperatur dalam K (screen_temp)	7 polygons
Kecepatan angin 10m dalam m/s (wind_speed)	7 polygons
Target	
Target	744 points

Pohon keputusan spasial yang dihasilkan terdiri dari 276 *node* daun dengan jarak ke sungai terdekat menjadi layer uji

pertama yang digunakan untuk pemisahan dataset spasial. Akurasi pohon keputusan spasial pada data training mencapai 87.69%. Beberapa aturan yang dihasilkan adalah sebagai berikut:

1. IF dist_river (km) > 3 AND 0 ≤ wind_speed (m/s) < 1 AND income_source = Forestry AND dist_road (km) ≤ 2.5 AND dist_city (km) > 14 THEN hotspot occurrence = T
2. IF dist_river (km) > 3 AND 0 ≤ wind_speed (m/s) < 1 AND income_source = Plantation AND land_cover = Shrubs AND dist_city (km) ≤ 7 THEN hotspot occurrence = F
3. IF dist_river (km) > 3 AND 1 ≤ wind_speed (m/s) < 2 AND 50 < population_density ≤ 100 AND land_cover = Shrubs AND income_source = Other_agriculture THEN hotspot occurrence = F
4. IF 1.5 < dist_river (km) ≤ 3 AND income_source = Forestry AND dist_city (km) > 14 THEN hotspot occurrence = T
5. IF dist_river (km) ≤ 1.5 AND land_cover = Bare_land AND dist_city (km) > 14 AND dist_road (km) ≤ 2.5 THEN hotspot occurrence = T

Aturan-aturan tersebut dapat digunakan untuk menentukan karakteristik dari area dimana titik api diprediksi akan muncul. Sebagai contoh, aturan 1 menyatakan bahwa pada area yang memiliki karakteristik berikut diprediksi adanya titik api:

1. Jarak ke sungai terdekat (dist_river) di atas 3 km
2. Kecepatan angin (wind_speed) di bawah 1 m/s
3. Sumber pendapatan penduduk yang tinggal di area tersebut (income_source) adalah hutan (Forestry)
4. Jarak ke jalan terdekat (dist_road) sedikitnya 2.5 km
5. Jarak ke pusat kota terdekat (dist_city) di atas 14 km

IV. PENUTUP

Teknik *data mining* meliputi klasifikasi, asosiasi dan clustering telah banyak digunakan dalam menganalisis data kebakaran hutan. Teknik klasifikasi menggunakan algoritme pohon keputusan merupakan salah satu metode yang dapat digunakan dalam membuat model prediksi resiko terjadinya titik api sebagai indikator dari kebakaran. Studi yang pernah dilakukan mengaplikasikan algoritme pohon keputusan konvensional ke dalam dataset kebakaran hutan di Provinsi Riau berdasarkan faktor fisik mencakup jaringan jalan, sungai, penutupan lahan dan kedalaman serta tipe lahan gambut. Algoritme pohon keputusan telah diperbaharui menghasilkan algoritme pohon keputusan spasial agar dapat diterapkan secara langsung ke dataset spasial. Algoritme tersebut telah diaplikasikan pada data spasial kebakaran hutan yang terdiri dari 10 layer fisik, sosial-ekonomi, dan cuaca. Model yang dihasilkan yaitu pohon keputusan spasial memiliki akurasi 87.69% pada data training. Model tersebut dapat digunakan untuk memprediksi kemunculan titik api pada daerah yang baru.

REFERENSI

- M. Darmawan, M. Aniya and S. Tsuyuki, "Forest fire hazard model using remote sensing and geographic information systems: toward understanding of land and forest degradation on lowland areas of east Kalimantan Indonesia", 2000.
- J. Boonyanuphap, "GIS-based method in developing wildfire risk model. a case study in Sasamba, East Kalimantan, Indonesia," Thesis. Graduate Program, Bogor Agricultural University, Indonesia, 2001. Unpublished.
- M. Hadi, "Pemodelan spasial kerawanan kebakaran di lahan gambut: studi kasus kabupaten Bengkalis, Provinsi Riau. Thesis. Graduate Program, Bogor Agricultural University, Indonesia, 2006. Unpublished.
- K. S. N. Prasad & S. Ramakrishna, "An autonomous forest fire detection system based on spatial *data mining* and fuzzy logic," *IJCSNS International Journal of Computer Science and Network Security*, vol.8 no.12, pp. 49-55, December 2008.
- D. Stojanova, P. Panov, A. Kobler, S. Džeroski, & K. Taškova, Learning to predict forest fires with different *data mining* techniques. Proc. Conference on *Data mining and Data Warehouses (SiKDD 2006)*, Ljubljana, Slovenia, 2006.
- P. Cortez and A. Morais, A *data mining* approach to predict forest fires using meteorological data.
- K. Angayarkkani & N. Radhakrishnan, "Efficient forest fire detection system: a spatial *data mining* and image processing based approach," *IJCSNS International Journal of Computer Science and Network Security*, vol.9 no.3, pp. 100-107, March 2009.
- HU Lin, ZHOU Goumin, & QIU Yun, "Application of Apriori Algorithm to the *Data mining* of the Wildfire," Proc. Sixth International Conference on Fuzzy Systems and Knowledge Discovery, IEEE, 2009, pp. 426-429.
- J. Han and M. Kamber, *Data mining: Concepts and Techniques*. Second Edition, The Morgan Kaufmann series in data management systems. Morgan Kaufmann, 2006.
- I. S. Sitanggang, and M. H. Ismail, "Classification model for hotspot occurrences using a decision tree method," *Journal of Geomatics*,

Natural Hazards and Risk, Taylor & Francis, vol.2: 2, pp. 111 – 121, 2011.

- I. S Sitanggang, R. Yaakob, N. Mustapha, A.N. Ainuddin, "Application of classification algorithms in *data mining* for hotspots occurrence prediction in Riau Province Indonesia," Journal of Theoretical and Applied Information Technology, vol. 43(2), 2012.
- I.S. Sitanggang, R. Yaakob, N. Mustapha and A.N. Ainuddin, "Predictive Models for Hotspots Occurrence using Decision Tree Algorithms and Logistic Regression," Journal of Applied Sciences, vol. 13(2). 2013.
- I.S Sitanggang, R. Yaakob, N. Mustapha and A.N. Ainuddin, "Classification Model for Hotspot Occurrences using Spatial Decision Tree Algorithm," Journal of Computer Science, Science Publications, vol. 9 (2), pp. 244-251, 2013.