

Optimasi Aturan Asosiasi Multidimensi menggunakan Algoritme Genetika untuk Klasifikasi Kemunculan Titik Panas

Multidimension Association Rule Optimization using Genetic Algorithm for hotspot occurrence Classification

NIA KURNIATI^{1*}, IMAS SUKAESIH SITANGGANG¹, IRMAN HERMADI¹

Abstrak

Penelitian ini menggunakan algoritme genetika untuk mengoptimalkan pembentukan aturan asosiasi yang dihasilkan dari algoritme *apriori*. Algoritme *apriori* diterapkan pada *dataset* kebakaran hutan dengan daerah penelitian di wilayah Rokan Hilir provinsi Riau. Aturan asosiasi diklasifikasi menggunakan algoritme CPAR (*Classification Based on Predictive Association Rules*) untuk mengetahui karakteristik wilayah yang berpotensi munculnya titik panas. Optimasi dalam algoritme genetika dilakukan melalui tahapan: skema pengkodean, evaluasi *fitness*, seleksi, pindah silang, mutasi. Hasil yang diperoleh dari penelitian ini ialah jumlah aturan dapat dikurangi. Jumlah aturan yang dihasilkan yaitu sebanyak 121 aturan pada generasi ke-300 hingga mencapai 108 aturan pada generasi ke-50. Hasil klasifikasi menunjukkan bahwa daerah yang berpotensi timbulnya titik panas yang terdapat pada generasi ke-50 ditemukan pada daerah yang memiliki curah hujan lebih besar dari atau sama dengan 3 mm per hari dan yang memiliki temperatur pada interval 297 Kelvin hingga 298 Kelvin dengan laplace akurasi sebesar 0.76. Pada generasi 150 ditemukan daerah yang memiliki temperatur yaitu pada interval 297 Kelvin hingga 298 Kelvin dengan laplace akurasi sebesar 0.57. Pada generasi 300 ditemukan pada daerah yang memiliki kecepatan angin yaitu pada interval 1 $m s^{-1}$ hingga 2 $m s^{-1}$ dengan laplace akurasi sebesar 0.70.

Kata kunci: algoritme *apriori*, algoritme CPAR, algoritme genetika, aturan asosiasi multidimensi, titik panas.

Abstract

This study used genetic algorithm to optimize *association rules* generated using the *Apriori* algorithm. *Apriori* algorithm was applied to forest fires multidimensional *dataset* with the study area in Rokan Hilir Riau province. *Association rules* were used to classify *hotspot* occurrence using the CPAR (*Prediction Classification based on Association Rule*) algorithm to determine the characteristics of areas are potential of *hotspot* occurrence. Optimization in genetic algorithms was done through these stages: *encoding* scheme, *fitness* evaluation, selection (*roulette wheel*), crossovers (*one point crossover*), and mutation. The results showed the proposed method successfully reduced the number of association rules. The number of rules generated as many as 121 in the generation of 300 to reach 108 rule in the 50th generation. Classification results show that the area has the potential emergence of hotspots were found on the generation of 50, found in areas where rainfall is greater than or equal to 3 mm per day and which has a temperature in the interval 297 Kelvin to 298 Kelvin by *Laplace accuracy* of 0.76. In the generation of 150, we found the area that have a temperature in the interval 297 Kelvin to 298 Kelvin by *Laplace accuracy* of 0.57. In the generation of 300, we found the area that have a wind speeds that is intervals of 1 $m s^{-1}$ up to 2 $m s^{-1}$ with *Laplace accuracy* of 0.70.

Keywords : *Apriori algorithm, association rule multidimension, CPAR algorithm, genetic algorithm, hotspot.*

PENDAHULUAN

Kebakaran hutan dan lahan menjadi suatu permasalahan yang serius di Indonesia. Kejadian kebakaran hutan di Indonesia meningkat dari tahun ke tahun. Kabupaten Rokan Hilir adalah salah satu wilayah yang sering mengalami kebakaran hutan dan lahan di Riau. Pada

¹Departemen Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor, Bogor

*Penulis korespondensi: Tel/Faks: 085242850385; Surel: nhia.imb@gmail.com

bulan Februari dan Maret tahun 2014 kebakaran yang terjadi telah menyebabkan kerugian hingga 20 triliun rupiah (Nugroho 2014). Menurut Syaufina (2008) kebakaran pada umumnya disebabkan oleh faktor manusia dan sebagian kecil karena faktor alam. Indikasi kebakaran hutan dan lahan dapat diketahui melalui titik panas yang terdeteksi di suatu lokasi tertentu pada saat tertentu (Syaufina 2008). Salah satu yang dapat diterapkan dalam upaya pencegahan kebakaran hutan ialah mengetahui karakteristik wilayah terjadinya titik panas dengan menggunakan teknik *data mining*. *Data mining* memiliki kemampuan untuk menemukan pola yang sering muncul di dalam *dataset* salah satunya yaitu menerapkan penambangan aturan asosiasi. Menurut Han *et al.* (2012) aturan asosiasi adalah pengetahuan untuk mengetahui keterkaitan dan bobot kepentingan objek yang satu dengan yang lainnya.

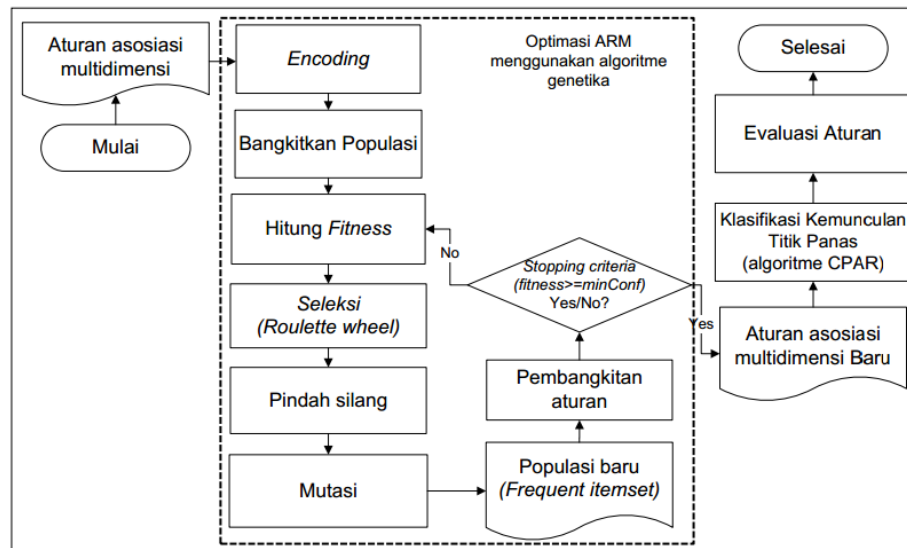
Aturan asosiasi multidimensi merupakan pengetahuan baru yang mulai banyak dikembangkan. Studi oleh Sitanggung (2013) menerapkan teknik *data mining* yaitu penambangan aturan asosiasi menggunakan algoritme apriori. Dengan menggunakan minimum *support* sebesar 10% dan minimum *confidence* sebesar 80% diperoleh sejumlah aturan untuk mengetahui faktor-faktor yang memengaruhi terjadinya titik panas sebagai indikator terjadinya peristiwa kebakaran hutan. Menurut Han *et al.* (2012) aturan asosiasi yang bersifat *multidimensional* disebut dengan *high dimensional pattern*. Aturan asosiasi spasial multidimensi cenderung memiliki banyak kombinasi yaitu lebih kompleks dan lebih bervariasi dari pada non-spasial. Aturan ini tidak hanya melibatkan satu predikat tapi juga dapat terdiri dari dua predikat atau lebih. Menurut Saggarr *et al.* (2004) Penambangan aturan asosiasi menghasilkan kompleksitas dalam aturan yang bersifat algoritmik. Salah satu kompleksitas yang ditimbulkan yaitu jumlah aturan tumbuh secara eksponensial berdasarkan jumlah *item*. Kompleksitas ini dapat ditangani dengan memangkas ruang pencarian dan menemukan aturan yang berkualitas. Kompleksitas yang ditemukan ini diperlukan suatu algoritme untuk menemukan kemungkinan-kemungkinan dalam memperoleh solusi yang optimal. Salah satu solusi optimal ialah menerapkan metode algoritme genetika.

Algoritme genetika merupakan metode penyelesaian solusi dari berbagai macam jenis permasalahan. Salah satu kasus permasalahan yang dapat diterapkan ialah yang berkaitan mengenai penambangan aturan asosiasi. Hsu *et al.* (2003) pertama kali menggabungkan algoritme aturan asosiasi dan algoritme genetika dalam melakukan pencarian secara efisien. Ghosh dan Nath (2004) menerapkan algoritme genetika dalam mengurangi aturan yang bersifat multiobjektif. Prakash *et al.* (2011) menghasilkan sejumlah *frequent itemset* dengan memperhatikan atribut diskrit dan interval dalam proses penemuan aturan dan berhasil mengurangi waktu proses komputasi Du *et al.* (2009) mengurangi aturan yang berlebihan dan meningkatkan efisiensi pencarian. Saggarr *et al.* (2004) mengoptimalkan kompleksitas aturan yang ditimbulkan dengan memperhatikan atribut negatif. Saxena *et al.* (2012) mengurangi proses iterasi dalam menghasilkan individu yang terbaik.

Tujuan dari penelitian ini adalah menerapkan algoritme genetika dengan membuat model optimasi pembangkitan aturan asosiasi berbasis apriori untuk mengoptimasi 1080 aturan asosiasi multidimensi yang dihasilkan oleh Sitanggung (2013). Diharapkan jumlah aturan yang dihasilkan lebih sedikit dan mengetahui model aturan yang dihasilkan. Aturan yang diperoleh kemudian dibentuk aturan klasifikasi menggunakan algoritme CPAR (*Classification Based on Predictive Association Rules*) untuk mengidentifikasi karakteristik wilayah yang berpotensi munculnya titik panas.

METODE

Penelitian ini mencoba mengurangi aturan asosiasi multidimensi yang telah dihasilkan oleh Sitanggung (2013) dengan menggunakan algoritme genetika. Penelitian ini terdiri atas 3 tahap, yaitu optimasi aturan asosiasi multidimensi menggunakan algoritme genetika, klasifikasi kemunculan titik panas (algoritme CPAR), dan evaluasi aturan. Dalam Goldberg (1989) optimasi dalam Algoritme Genetika dilakukan melalui tahapan : skema pengkodean (*encoding*), evaluasi *fitness*, seleksi, pindah silang, dan mutasi. Alur penelitian pada Gambar 1.



Gambar 5 Alur penelitian untuk optimasi aturan asosiasi menggunakan Algoritme Genetika

Data Lokasi

Data spasial yang digunakan mengacu pada data penelitian sebelumnya (Sitanggang 2013) yaitu *dataset* kebakaran hutan wilayah Kabupaten Rokan Hilir provinsi Riau tahun 2008. Data ini meliputi kondisi geografis, ekonomi, sosial dan cuaca. Adapun data dan sumber data yang diperoleh diuraikan pada Tabel 1.

Tabel 1 Data dan sumber data (Sitanggang 2013)

Data	Tahun	Sumber
Titik panas	2008	FIRMS MODIS Fire/Titik panas, NASA/University of Maryland
Suhu harian maksimal, curah hujan harian, dan kecepatan angin	2008	Badan Meterologi Klimatologi dan Geofisika (BMKG)
Sungai, jalan, dan tutupan lahan	2008	Badan Koordinasi Survey dan Pemetaan Nasional (BAKOSURTANAL)
Tipe lahan gambut, kedalaman dan lahan gambut		Weatland International
Sumber pendapatan		BPS (Badan Pusat Statistik)

Optimasi aturan asosiasi multidimensi menggunakan Algoritme Genetika

Penelitian ini bertujuan untuk mengurangi aturan asosiasi multidimensi yang telah dihasilkan oleh Sitanggang (2013) dengan menggunakan algoritme genetika. Jumlah aturan asosiasi yang diperoleh sebanyak 1080 aturan yang terdiri dari fitur spasial.

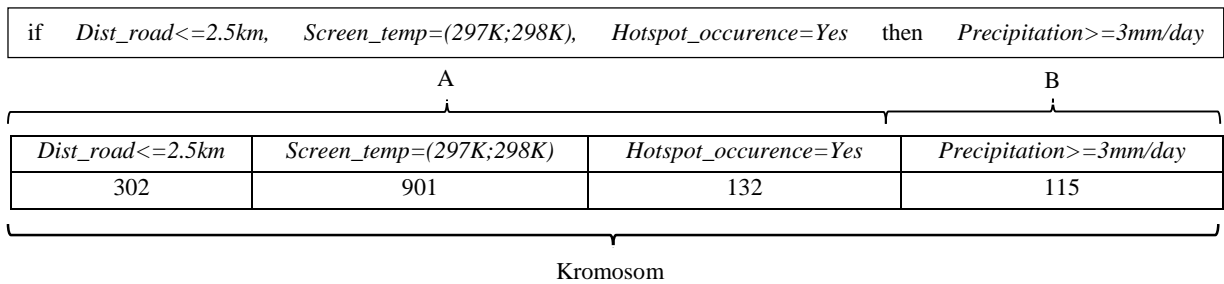
Pembangkitan Populasi

Tahap awal yang dilakukan dalam proses algoritme genetika ialah membangkitkan populasi sebanyak jumlah kromosom. Pada tahap ini Jumlah ukuran populasi yang dibangkitkan ialah sebanyak 1080 kromosom.

Skema pengkodean (*Encoding*)

Skema pengkodean dilakukan melalui pendekatan Michigan. Menurut Du *et al.* (2009) aturan dikodekan sebagai nomor baku. Setiap aturan dikodekan ke dalam kromosom (Gupta dan Sikka 2013). Representasi kromosom dapat dilakukan menggunakan penomoran, pohon, *bit*, *array*, daftar atau dalam bentuk lainnya (Indira dan Kanmani 2012). Pengkodean dilakukan disebabkan jumlah *string* yang cukup panjang dan tidak efisien. Pada tahap ini masing-masing variabel *string* diterjemahkan berdasarkan ukuran dan jumlah data. Setiap 1 (satu) *string* membawa 3 digit angka di mana setiap gen ditempatkan berdasarkan urutan masing-masing. Pada Gambar 2 menunjukkan pembentukan kromosom yang dikodekan di mana *A* adalah *antecedent* (bagian jika), *B* adalah *consequent* (bagian maka).

Contoh *rule* ke-78 :



Gambar 2 Pembentukan kromosom

Evaluasi *fitness*

Evaluasi nilai *fitness* adalah nilai yang menyatakan baik tidaknya suatu solusi (kromosom). Nilai *fitness* yang paling tinggi menjadi tolak ukur optimal suatu solusi. Semakin optimal nilai *fitness* yang dihasilkan maka kualitas kromosom yang memenuhi nilai *fitness* tersebut dinyatakan semakin baik. Menurut Du *et al.* (2009) *support* dan *confidence* merupakan parameter dari nilai *fitness*. Variabel *a* dan variabel *b* merupakan koefisien bobot untuk mengontrol nilai *fitness*. Koefisien bobot dinyatakan sebagai $0 < a \leq 1$ dan $0 < b \leq 1$ di mana $a = 0.5$ dan $b = 0.5$. Nilai *fitness* diberikan pada persamaan 1 (Du *et al.* 2009):

$$fitness(r) = a \times sup(r) + b \times conf(r) \quad (1)$$

dengan,

sup : *Support* dari aturan *r* *a* : Koefisien bobot nilai *support*
conf : *Confidence* dari aturan *r* *b* : Koefisien bobot nilai *confidence*

Seleksi

Proses seleksi menggunakan roulette wheel yaitu menyeleksi populasi baru dengan melakukan distribusi probabilitas berdasarkan hasil probabilitas yang terpilih. Kromosom yang terpilih dievaluasi dengan menggunakan nilai *fitness* kemudian diseleksi untuk dijadikan induk (Goldberg 1989).

Pindah Silang

Pindah silang (*crossover*) adalah operator algoritme genetika yang dilakukan pada sepasang kromosom induk untuk menghasilkan dua kromosom anak dengan menukarkan beberapa elemen masing-masing kromosom. Parameter pindah silang yang baik ialah 0.6-1.0. (Goldberg 1989). Tahap ini dilakukan *one cut point crossover* dengan parameter sebesar 0.6.

Mutasi

Tahap mutasi dilakukan dengan menentukan nilai awal secara acak dengan menggunakan parameter peluang mutasi (Goldberg 1989). Talbi (2009) menganjurkan mutasi yang cukup kecil yaitu antara 0.001-0.01. Proses mutasi dilakukan dengan memilih dua gen secara acak kemudian menukar (*swap*) dua gen yang terpilih dengan parameter mutasi. Mutasi yaitu sebesar 0.001 dikali dengan banyaknya domain variabel yang digunakan.

Kriteria berhenti

Kriteria berhenti (*stopping criteria*) ialah membatasi jumlah generasi dengan berdasarkan syarat fungsi *fitness* (Engelbrecht 2007). *Stopping criteria* terjadi ketika nilai *fitness* lebih besar atau sama dengan minimum *confidence* (Jaiswal dan Dubey 2013).

Pembangkitan aturan

Populasi baru yang dihasilkan oleh algoritme genetika dibangkitkan melalui algoritme apriori. Apriori adalah algoritme untuk mencari *frequent itemset* dengan menggunakan *teknik association rule* (Agrawal dan Srikant 1994).

Penambangan aturan asosiasi

Penambangan aturan asosiasi merupakan implikasi dari bentuk A ke B , di mana A dan B adalah 2 himpunan *itemset* (Han *et al.* 2012). Aturan asosiasi bekerja pada *dataset*. Rumus *support* diberikan dalam persamaan 2 (Han *et al.* 2012):

$$\text{Support}(A \Rightarrow B) = \frac{\text{count}(A \cup B)}{D} \quad (2)$$

Rumus *confidence* diberikan dalam persamaan 3 (Han *et al.* 2012):

$$\text{Confidence}(A \Rightarrow B) = \frac{\text{count}(A \cup B)}{\text{count}(A)} \quad (3)$$

dengan,

A : *Antecedent* aturan (bagian jika),

B : *Consequent* aturan (bagian maka),

D : *Dataset (transaction)*

Klasifikasi asosiatif (*Associative Classification*)

Klasifikasi asosiatif merupakan penemuan aturan asosiasi yang berupa aturan jika-maka di mana yang terdapat dalam set aturan seperti $X \Rightarrow Y$, Y dijadikan sebagai atribut kelas (label kelas) (Balaji dan Rao 2013).

Algoritme CPAR

CPAR (*Classification based on Predictive Association Rule*) adalah algoritme yang berawal dari FOIL (*First Order Inductive Learner*) yang menggunakan algoritme *greedy* dalam membangkitkan aturan. Oleh Yin dan Han (2003) parameter *Total Weight Threshold* (TWT)=0.05, *minimum gain*=0.7, dan *decay factor*=2/3. Formula untuk TWT dihitung dengan persamaan 4 berikut (Herwanto dan Arymurthy (2010); Yin dan Han 2003):

$$\text{TWT} = \text{Bobot sampel positif} / \text{bobot sampel negatif} * 0.05 \quad (4)$$

FOIL *Gain* digunakan untuk mengukur informasi dari penambahan *literal* ke *current rule*. Pada persamaan 5 *Gain* dihitung dengan formula (Yin dan Han 2003):

$$\text{Gain}(p) = |P^*| \left[\log \frac{|P^*|}{|P^*| + |N^*|} - \log \frac{|P|}{|P| + |N|} \right] \quad (5)$$

dengan,

$|P|$: banyaknya objek yang mengandung sampel positif

$|N|$: banyaknya objek yang mengandung sampel negatif yang memenuhi set aturan

$|P^*|$: jumlah sampel yang mengandung sampel positif

$|N^*|$: jumlah sampel yang mengandung sampel negatif yang memenuhi set aturan baru.

Formula untuk LGT dihitung dengan persamaan 6 berikut (Yin dan Han 2003):

$$\text{LGT (Local Gain Threshold)} = \text{Gain terbesar} * \text{Gain similarity Ratio}(\%) \quad (6)$$

Laplace accuracy (LA) yaitu evaluasi aturan untuk memprediksi dan mengetahui kekuatan dari suatu aturan (Yin dan Han 2003). Formula untuk *laplace akurasi* (LA) dihitung dengan persamaan (7) berikut (Yin dan Han 2003):

$$\text{laplace akurasi (LA)} = (n_c+1)/(n_{tot}+k) \tag{7}$$

dengan,

k : jumlah kelas

n_{tot} : jumlah total aturan dari set aturan yang terpilih

n_c : jumlah aturan terpilih yang diprediksi

HASIL DAN PEMBAHASAN

Optimasi menggunakan Algoritme Genetika

Optimasi menggunakan algoritme genetika dimulai dengan membangkitkan populasi awal secara acak. Populasi awal yaitu sebanyak 1080 *record* aturan.

Pengujian Hasil Optimasi

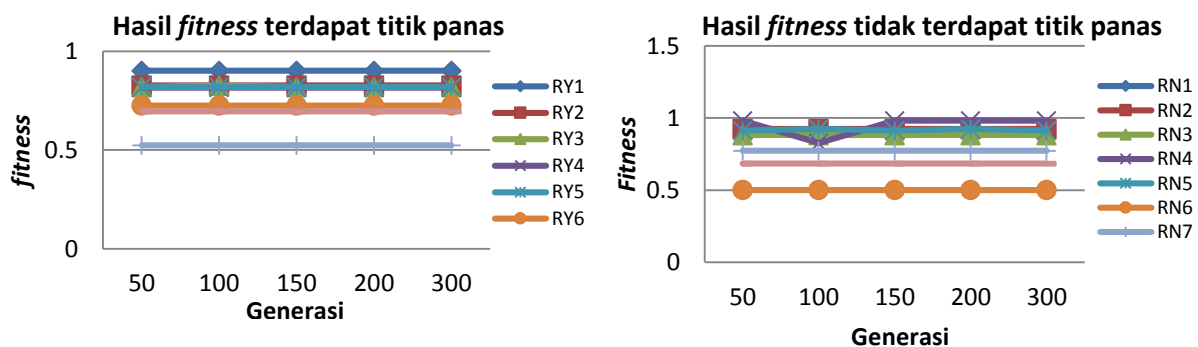
Berdasarkan hasil pengujian bahwa aturan dapat dikurangi. Pada Tabel 2 diperoleh jumlah aturan yang memiliki rata-rata nilai *fitness* paling tinggi. Nilai *fitness* diperoleh dari nilai *support* dan nilai *confidence*. Dapat dilihat pada Tabel 2. Jumlah aturan asosiasi dan hasil rata-rata nilai *fitness* dari proses algoritme genetika.

Tabel 2 Jumlah aturan asosiasi dan nilai *fitness* dari proses algoritme genetika

Generasi	Terdapat titik panas		Tidak terdapat titik panas		Jumlah rule
	Record	Fitness	Record	Fitness	
50	38	0.827	70	0.877	108
100	36	0.824	76	0.851	112
150	40	0.831	78	0.872	118
200	42	0.831	74	0.832	116
300	44	0.833	77	0.873	121

Aturan yang mempunyai kualitas yang tinggi dapat digunakan untuk memprediksi data (Yin dan Han 2003). Pada Tabel 2 memperlihatkan bahwa pada tiap generasi diperoleh nilai *fitness* yang tertinggi terdapat titik panas yaitu pada generasi ke-300 dengan rata-rata nilai *fitness* sebesar 0.833. Sedangkan nilai *fitness* yang tertinggi tidak terdapat titik panas yaitu pada generasi ke-50 dengan rata-rata nilai *fitness* sebesar 0.877. Pada Gambar 3 diperoleh grafik garis hasil *fitness* yang terdapat titik panas dan hasil *fitness* yang tidak terdapat titik panas.

Grafik garis untuk *fitness* yang terdapat titik panas pada Gambar 3 memperlihatkan bahwa nilai *fitness* yang memenuhi syarat kriteria untuk masing-masing aturan ditemukan pada aturan dengan kode RY1,RY2,RY3,RY4,dan R58 dengan rentang nilai *fitness* sebesar 0.8 hingga 0.9. Grafik garis untuk hasil *fitness* yang tidak terdapat titik panas pada Gambar 3 memperlihatkan bahwa nilai *fitness* yang memenuhi syarat kriteria untuk masing-masing aturan ditemukan pada aturan dengan kode RN1,RN2,RN3,RN4,dan RN5 dengan rentang nilai *fitness* 0.8 hingga 0.9. Aturan ini dinyatakan aturan yang berkualitas karena memenuhi syarat kriteria.



Gambar 3 Hasil *fitness* terdapat titik panas dan hasil *fitness* tidak terdapat titik panas

Keterangan :

RY1= $precipitation \geq 3mm/day \Rightarrow hotspot_occurrence = Yes$
 RY2= $precipitation \geq 3mm/day$ dan $school \leq 0.1 \Rightarrow hotspot_occurrence = Yes$
 RY3= $precipitation \geq 3mm/day$ dan $screen_temp(297K;298K) \Rightarrow hotspot_occurrence = Yes$
 RY4= $school \leq 0.1 \Rightarrow hotspot_occurrence = Yes$
 RY5= $screen_temp(297K;298K) \Rightarrow hotspot_occurrence = Yes$
 RY6= $population \leq 50$ dan $screen_temp(297K;298K) \Rightarrow hotspot_occurrence = Yes$
 RY7= $precipitation \geq 3mm/day$ dan $peatland_depth = non_peatland \Rightarrow hotspot_occurrence = Yes$
 RY8= $screen_temp(297K;298K)$ dan $school \leq 0.1 \Rightarrow hotspot_occurrence = Yes$

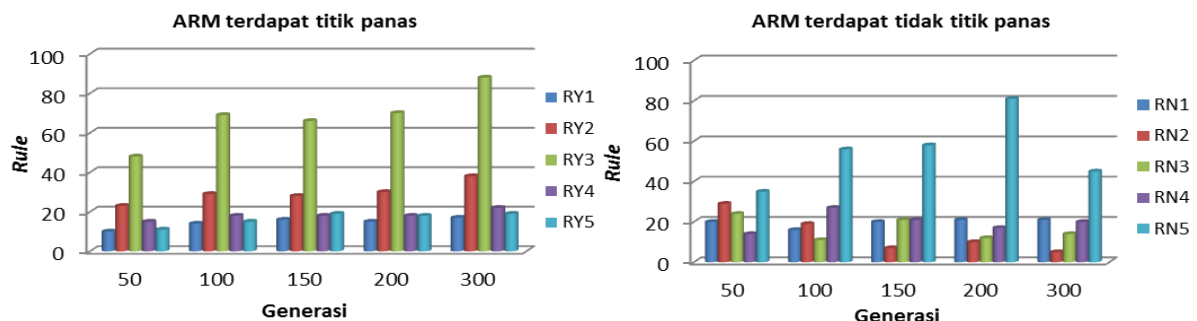
RN1= $dist_river \leq 1.5km$ dan $peatland_depth = non_peatland \Rightarrow hotspot_occurrence = No$
 RN2= $peatland_depth = non_peatland$ dan $peatland_type = non_peatland \Rightarrow hotspot_occurrence = No$
 RN3= $population \leq 50$ dan $school \leq 0.1 \Rightarrow hotspot_occurrence = No$
 RN4= $precipitation \geq 3mm/day$ dan $school \leq 0.1 \Rightarrow hotspot_occurrence = No$
 RN5= $precipitation \geq 3mm/day$ dan $peatland_type = non_peatland \Rightarrow hotspot_occurrence = No$
 RN6= $income_source = Trading_restaurant$ dan $peatland_depth = non_peatland \Rightarrow hotspot_occurrence = No$
 RN7= $population \leq 50$ dan $peatland_depth = non_peatland \Rightarrow hotspot_occurrence = No$
 RN8= $precipitation \geq 3mm/day$ dan $peatland_type = non_peatland \Rightarrow hotspot_occurrence = No$

Berikut Tabel 3 contoh 2 record aturan asosiasi yang terdapat variabel titik panas dan contoh 2 record tidak terdapat variabel titik panas pada tiap generasi.

Tabel 3 Aturan asosiasi tiap generasi

Generasi	Aturan asosiasi	Fitness
50	$precipitation \geq 3mm/day \Rightarrow hotspot_occurrence = Yes$	0.901
	$screen_temp(297K;298K) \Rightarrow hotspot_occurrence = Yes$	0.818
	$peatland_depth = non_peatland$ dan $peatland_type = non_peatland \Rightarrow hotspot_occurrence = No$	0.924
	$dist_river \leq 1.5km$ dan $peatland_depth = non_peatland \Rightarrow hotspot_occurrence = No$	0.887
100	$precipitation \geq 3mm/day \Rightarrow hotspot_occurrence = Yes$	0.901
	$school \leq 0.1 \Rightarrow hotspot_occurrence = Yes$	0.816
	$peatland_depth = non_peatland$ dan $peatland_type = non_peatland \Rightarrow hotspot_occurrence = No$	0.924
	$precipitation \geq 3mm/day, school \leq 0.1 \Rightarrow hotspot_occurrence = No$	0.882
150	$precipitation \geq 3mm/day$ dan $school \leq 0.1 \Rightarrow hotspot_occurrence = Yes$	0.826
	$precipitation \geq 3mm/day$ dan $screen_temp(297K;298K) \Rightarrow hotspot_occurrence = Yes$	0.818
	$peatland_depth = non_peatland$ dan $peatland_type = non_peatland \Rightarrow hotspot_occurrence = No$	0.924
	$precipitation \geq 3mm/day$ dan $peatland_type = non_peatland \Rightarrow hotspot_occurrence = No$	0.913
200	$precipitation \geq 3mm/day \Rightarrow hotspot_occurrence = Yes$	0.901
	$screen_temp(297K;298K) \Rightarrow hotspot_occurrence = Yes$	0.818
	$peatland_depth = non_peatland$ dan $peatland_type = non_peatland \Rightarrow hotspot_occurrence = No$	0.924
	$precipitation \geq 3mm/day$ dan $peatland_type = non_peatland \Rightarrow hotspot_occurrence = No$	0.924
300	$precipitation \geq 3mm/day$ dan $school \leq 0.1 \Rightarrow hotspot_occurrence = Yes$	0.826
	$screen_temp(297K;298K) \Rightarrow hotspot_occurrence = Yes$	0.818
	$precipitation \geq 3mm/day$ dan $peatland_type = non_peatland \Rightarrow hotspot_occurrence = No$	0.913
	$population \leq 50$ dan $school \leq 0.1 \Rightarrow hotspot_occurrence = No$	0.882

Hasil fitness dari tiap generasi yang terdapat titik panas dan yang tidak terdapat titik panas menghasilkan sejumlah aturan di mana diperlihatkan pada Gambar 4.



Gambar 4 ARM terdapat titik panas dan ARM tidak terdapat titik panas

Gambar 4 memperlihatkan diperoleh aturan yang terdapat titik panas yang sering muncul yaitu aturan dengan kode RY3, yang menjelaskan jika $precipitation \geq 3mm/day$ dan $screen_temp(297K;298K)$ maka $hotspot_occurrence = Yes$. Aturan yang tidak terdapat titik panas

yang sering muncul adalah aturan dengan kode RN5, yang menjelaskan jika $precipitation \geq 3mm/day$ dan $peatland_type = non_peatland$ maka $hotspot_occurrence = No$.

Klasifikasi kemunculan titik panas menggunakan algoritme CPAR

Aturan yang telah diperoleh diklasifikasi untuk sampel data. Pelatihan sampel data dilakukan sebesar 60% untuk data latih dan 40% data uji. Data latih dan data uji memuat informasi tentang jarak ke pusat kota terdekat, jarak ke jalan terdekat, jarak ke sungai terdekat, sumber pendapatan, tutupan lahan, kerapatan populasi, curah hujan, sekolah, suhu, kecepatan angin, tipe lahan gambut, kedalaman lahan gambut, kemunculan titik panas. Tabel 4 menunjukkan pembagian sampel data dari tiap generasi.

Tabel 4 Pembagian sampel data dari tiap generasi

Generasi	Jumlah Aturan	Data latih	Data uji
50	108	65	43
150	118	72	47
300	121	73	48

Pembagian sampel positif dan sampel negatif dilakukan terhadap data latih yang terdapat pada generasi 50, 150, dan 300. Berikut Tabel 5 jumlah sampel data latih dari tiap generasi.

Tabel 5 Jumlah sampel data latih dari tiap generasi

Generasi	Jumlah data latih	Sampel Positif	Sampel negatif
50	65	28	37
150	72	28	44
300	73	29	44

Himpunan aturan terdiri dari 2 label kelas yaitu kelas positif dengan keterangan ' $hotspot_occurrence = Yes$ ' (terdapat titik panas) dan kelas negatif dengan keterangan ' $hotspot_occurrence = No$ ' (tidak terdapat titik panas). Setiap *record* diberikan bobot=1. Bobot total diproses secara terus menerus sampai bobot total lebih kecil dari TWT. TWT dihitung berdasarkan formula pada persamaan 4. Tabel 6 menunjukkan TWT untuk setiap generasi.

Tabel 6 Total Weight Threshold (TWT) untuk setiap generasi

Generasi	Bobot total positif (<i>record</i>)	Bobot total negatif (<i>record</i>)	TWT	
			Bobot total sampel positif	Bobot total sampel negatif
50	28	37	1.40	1.85
150	28	44	1.40	2.20
300	29	44	1.45	2.20

Pembentukan Gain

Atribut yang terpilih ialah atribut yang menghasilkan *Gain* terbesar (*Best Gain*) dan telah memenuhi syarat LGT. LGT dihitung berdasarkan persamaan 6. Total bobot untuk setiap kategori yang memiliki *Gain* terbesar diuji dengan *gain similarity ratio* sebesar 60%. Tabel 7 *Gain* terbesar untuk sampel positif dan sampel negatif dengan LGT (bobot=1) setiap generasi.

Tabel 7 Hasil *Gain* terbesar dengan LGT (bobot=1) untuk setiap generasi

Sampel	Generasi	Keterangan atribut	<i>Gain</i> terbesar	LGT
Positif	50	<i>Screen_temp</i> (297K;298K)	3.84	2.30
		<i>Precipitation</i> $\geq 3mm/day$	2.97	
		<i>Population</i> ≤ 50	2.73	
	150	<i>Precipitation</i> $\geq 3mm/day$	5.52	3.31
		<i>Screen_temp</i> (297K;298K)	4.13	
	300	<i>Precipitation</i> $\geq 3mm/day$	4.56	2.73
<i>Wind_speed</i> (1m/s;2m/s)		3.81		
		<i>School</i> ≤ 0.1	3.71	

Sampel	Generasi	Keterangan atribut	Gain terbesar	LGT
Negatif	50	<i>Screen_temp(297K;298K)</i>	3.84	2.30
		<i>Precipitation</i> \geq <i>3mm/day</i>	2.97	
		<i>Population</i> \leq <i>50</i>	2.73	
	150	<i>Precipitation</i> \geq <i>3mm/day</i>	5.52	3.31
		<i>Screen_temp(297K;298K)</i>	4.13	
		<i>Precipitation</i> \geq <i>3mm/day</i>	4.56	
300	<i>Wind_speed(1m/s;2m/s)</i>	3.81	2.73	
	<i>School</i> \leq <i>0.1</i>	3.71		

Tabel 7 memperlihatkan hasil *Gain* terbesar dari setiap sampel positif dan sampel negatif menghasilkan nilai yang sama. Dengan demikian berdasarkan Tabel 7 menjelaskan bahwa atribut yang memiliki *Gain* terbesar dari LGT pada generasi 50 adalah atribut *Screen_temp(297K;298K)* dengan *Gain* sebesar 3.84. Generasi 150 adalah atribut *Precipitation* \geq *3mm/day* dengan *Gain* sebesar 5.52. Untuk generasi 300 adalah atribut *Precipitation* \geq *3mm/day* dengan *Gain* sebesar 4.56. Selanjutnya bobot yang terdapat pada sampel positif dan negatif direvisi menggunakan *decay factor*. Apabila setiap bobot masih lebih besar dari total bobot (TWT) maka setiap atribut diturunkan hingga lebih kecil dari total bobot. Seluruh total bobot pada setiap sampel positif dan sampel negatif dihitung berdasarkan persamaan 4. Kemudian menghitung *Gain* berdasarkan formula pada persamaan 5. Tahap ini secara terus menerus dilakukan hingga seluruh nilai *Gain* yang diperoleh tersebut lebih kecil dari *global minimum threshold*. Tabel 8 hasil *Gain* terbesar dengan LGT (bobot=0.66).

Tabel 8 Hasil *Gain* terbesar dengan LGT (bobot=0.66) untuk setiap generasi

Sampel	Generasi	Keterangan atribut	Gain terbesar	LGT
Positif	50	<i>Screen_temp(297K;298K)</i>	1.21	0.72
		<i>Population</i> \leq <i>50</i>	1	
		<i>Wind_speed(1 m s⁻¹;2 m s⁻¹)</i>	0.84	
	150	<i>Screen_temp(297K;298K)</i>	1.46	0.96
		<i>Wind_speed(1 m s⁻¹;2 m s⁻¹)</i>	2.03	
		<i>School</i> \leq <i>0.1</i>	1.35	
Negatif	50	<i>Precipitation</i> \geq <i>3mm/day</i>	5.18	3.18
		<i>Population</i> \leq <i>50</i>	3.65	
		<i>Precipitation</i> \geq <i>3mm/day</i>	9.62	
	150	<i>Precipitation</i> \geq <i>3mm/day</i>	4.56	2.73
		<i>Wind_speed(1 m s⁻¹;2 m s⁻¹)</i>	3.81	
		<i>School</i> \leq <i>0.1</i>	3.71	

Pada Tabel 8 menunjukkan bahwa *Gain* yang diperoleh untuk sampel positif dan sampel negatif pada setiap generasi masih lebih besar dari *global minimum* yaitu 0.7, sehingga setiap atribut direvisi kembali dengan bobot sebesar 0.33. Tabel 9 hasil *Gain* terbesar dengan LGT.

Tabel 9 Hasil *Gain* terbesar dengan LGT (bobot=0.33)

Sampel	Generasi	Keterangan atribut	Gain terbesar	LGT
Positif	50	<i>Wind_speed(1m/s;2m/s)</i>	0.55	0.33
	150	<i>Screen_temp(297K;298K)</i>	1.46	8.70
		<i>Precipitation</i> \geq <i>3mm/day</i>	7.74	
Negatif	50	<i>Screen_temp(297K;298K)</i>	7.16	4.64
		<i>Population</i> \leq <i>50</i>	4.68	
	150	<i>Precipitation</i> \geq <i>3mm/day</i>	14.50	8.70
	300	<i>Precipitation</i> \geq <i>3mm/day</i>	13.78	8.26

Tabel 9 memperlihatkan bahwa untuk sampel positif pada generasi ke 50 ditemukan atribut yang berisi *Population* \leq *50* dengan *Gain* sebesar 0.55 yang berarti bahwa nilai *Gain* yang diperoleh lebih kecil dari *global minimum* sehingga proses dihentikan. Pada generasi 150

untuk sampel positif menghasilkan *Gain* yang lebih besar dari *global minimum* sehingga direvisi kembali dengan bobot sebesar 0.11. Pada generasi 300 untuk sampel positif tidak ditulis karena *Gain* dihasilkan kurang dari *global minimum* sehingga proses revisi dihentikan. Pada Tabel 9 memperlihatkan bahwa *Gain* yang dihasilkan pada setiap generasi masih lebih besar dari *global minimum* sehingga direvisi kembali dengan bobot sebesar 0.11. Pada akhir proses semua aturan dihitung menggunakan formula laplace akurasi (LA) pada persamaan 7. Tabel 10 hasil laplace akurasi setiap aturan dengan *Gain similarity ratio* sebesar 60%.

Tabel 10 Aturan dan hasil laplace akurasi (LA) dengan *Gain similarity ratio* 60%

Generasi	Aturan	L.A
50	If <i>Population</i> ≤ 50 Then <i>hotspot_occurence</i> = Yes	0.38
	If <i>Precipitation</i> ≥ 3mm/day dan <i>Screen_temp</i> (297K;298K) Then <i>hotspot_occurence</i> = Yes	0.76
	If <i>Population</i> ≤ 50 Then <i>hotspot_occurence</i> = No	0.61
	If <i>Precipitation</i> ≥ 3mm/day dan <i>Screen_temp</i> (297K;298K) Then <i>hotspot_occurence</i> = No	0.21
150	If <i>Precipitation</i> ≥ 3mm/day Then <i>hotspot_occurence</i> = Yes	0.52
	If <i>Screen_temp</i> (297K;298K) Then <i>hotspot_occurence</i> = Yes	0.57
	If <i>Precipitation</i> ≥ 3mm/day Then <i>hotspot_occurence</i> = No	0.47
	If <i>Screen_temp</i> (297K;298K) Then <i>hotspot_occurence</i> = No	0.42
300	If <i>Wind_speed</i> (1m/s;2m/s) Then <i>hotspot_occurence</i> = Yes	0.70
	If <i>School</i> ≤ 0.1 Then <i>hotspot_occurence</i> = Yes	0.58
	If <i>Wind_speed</i> (1m/s;2m/s) Then <i>hotspot_occurence</i> = No	0.30
	If <i>School</i> ≤ 0.1 Then <i>hotspot_occurence</i> = No	0.41

Evaluasi aturan

Pada Tabel 10 untuk kelas positif memperlihatkan pada generasi 50 diperoleh aturan “If *Precipitation* ≥ 3mm/day dan *Screen_temp*(297K;298K) Then *hotspot_occurence* = Yes dengan hasil laplace akurasi paling tinggi yaitu sebesar 0.76. Generasi 300 diperoleh aturan If *Wind_speed*(1m/s;2m/s) Then *hotspot_occurence* = Yes dengan hasil laplace akurasi paling tinggi yaitu 0.70. Generasi 150 diperoleh aturan If *Screen_temp*(297K;298K) Then *hotspot_occurence* = Yes menghasilkan laplace akurasi paling tinggi yaitu 0.57. Berdasarkan aturan yang dihasilkan pada Tabel 3 dan hasil *fitness* yang diperlihatkan pada Tabel 2 membuktikan bahwa daerah yang terdapat titik panas dari aturan yang ditemukan dapat dinyatakan bahwa daerah tersebut berpotensi memiliki kemungkinan dapat menimbulkan titik panas berdasarkan hasil dari persentase LA yang diperoleh. Hasil persentase menunjukkan bahwa daerah yang memiliki peluang munculnya titik panas yang paling tinggi ditemukan pada daerah yang memiliki curah hujan lebih besar dari atau sama dengan 3 mm per hari dengan temperatur pada interval 297 Kelvin hingga 298 Kelvin dengan LA sebesar 76%.

SIMPULAN

Optimasi terhadap aturan asosiasi multidimensi dapat diterapkan dengan menggunakan algoritme genetika. Melalui algoritme genetika jumlah aturan yang dihasilkan dapat dikurangi yaitu sebanyak 121 hingga 108 aturan asosiasi. Rata-rata nilai *fitness* yang diperoleh pada tiap generasi yaitu sebesar 0.8 hingga 0.9. Hasil klasifikasi menunjukkan bahwa daerah yang berpotensi timbulnya titik panas yang terdapat pada generasi ke-50 ditemukan pada daerah yang memiliki curah hujan lebih besar dari atau sama dengan 3 mm per hari dan yang memiliki temperatur pada interval 297 Kelvin hingga 298 Kelvin dengan laplace akurasi sebesar 0.76. Generasi 150 ditemukan daerah yang memiliki temperatur yaitu pada interval 297 Kelvin hingga 298 Kelvin dengan laplace akurasi sebesar 0.57. Generasi 300 ditemukan pada daerah yang memiliki kecepatan angin yaitu pada interval 1 m s⁻¹ hingga 2 m s⁻¹ dengan laplace akurasi sebesar 0.70.

DAFTAR PUSTAKA

Agrawal R, Srikant R. 1994. Fast algorithms for mining association rules, *Proceedings of the VLDB Conference*. 487-489.

- Balaji BV, Rao VV. 2013. Improved classification based association rule mining, *International Journal of Advanced Research in Computer and Communication Engineering*. 2(5):2211-2221.
- Du F, Rao N, Guo J, Yuan Z, Wang R. 2009. Mining gene network by combined association rules and genetic algorithm, *IEEE Journal*. 581-585.
- Engelbrecht AP. 2007. *Computational Intelligence*. Second Edition. University of Pretoria.
- Ghosh A., Nath B. 2004. *Elsevier*. 123-133.
- Golberg DE. 1989. *Genetic Algorithms in Search, Optimization & Machine Learning*. Boston (US): Addison-Wesley.
- Gupta MK, Sikka G. 2013. Association rules extraction using multi-objective feature of genetic algorithm, *Proceedings of the World Congress on Engineering and Computer Science WCECS 2013*. Vol.2
- Han J., Kamber M., Pei J. 2012. *Data Mining: Concepts and Techniques*. San Francisco,(US):Morgan–Kauffman Publisher.
- Herwanto, Arymurthy AM. 2010. Penerapan sistem data mining untuk diagnosis penyakit kanker payudara menggunakan classification based on association algorithm. *JUTI*. 8(2).
- Hsu P. L., Lai R., Chiu C. C., Hsu C. I. 2003. The hybrid of association rule algorithms and genetic algorithms for tree induction: an example of predicting the student course performance. *Expert System with Application. Journal elsevier. Pergamon*.51-62.
- Indira K., Kanmani S. 2012. Performance analysis of genetic algorithm for mining association rules. *International Journal of Computer Science(IJCSI)*. 9(1):368-376.
- Jaiswal A, Dubey G. 2013. Identifying best association rules and their optimization using genetic algorithm. *International Journal of Emerging Science and Engineering*. 1(7).
- Nugroho SP. 2014. Kebakaran Hutan, Menimbulkan Kerugian Ekonomi Terparah. [Internet]. [diunduh 2015 Agustus 12]. Tersedia pada: <http://www.mongabay.co.id/2014/05/03/kebakaran-hutan-menimbulkan-kerugian-ekonomi-terparah/>.
- Prakash RV, Govardan, SarmaSSVN. 2011. Mining frequent itemsets from large data sets using genetic algorithm. *IJCA*. Warangal, India. 38-43.
- Saxena MR, Shrivastava S, Mathur A. 2012. Association rules mining using modified genetic algorithm. *International Journal of Scientific Engineering and Technology*. 1(4).
- Saggat M, Agrawal AK, Lad A. 2004. Optimization of association rule mining using improved genetic algorithms. *IEEE Indian Institute of Information Technology*. India.
- Sitanggang IS. 2013. Spatial multidimensional association rules mining in forest fire data. *Journal of Data Analysis and Information Processing*. 1:90-96.
- Syaufina L. 2008. *Kebakaran Hutan dan Lahan di Indonesia: Perilaku Api, Penyebab dan Dampak Kebakaran*. Malang (ID): Bayumedia Publishing.
- Talbi, EG. 2009. *Metaheuristik: From Design To Implementation*, New Jersey :JohnWiley&Son.
- Yin X, Han J. 2003. CPAR: Classification based on predictive association rules. *Proceedings of Siam International*