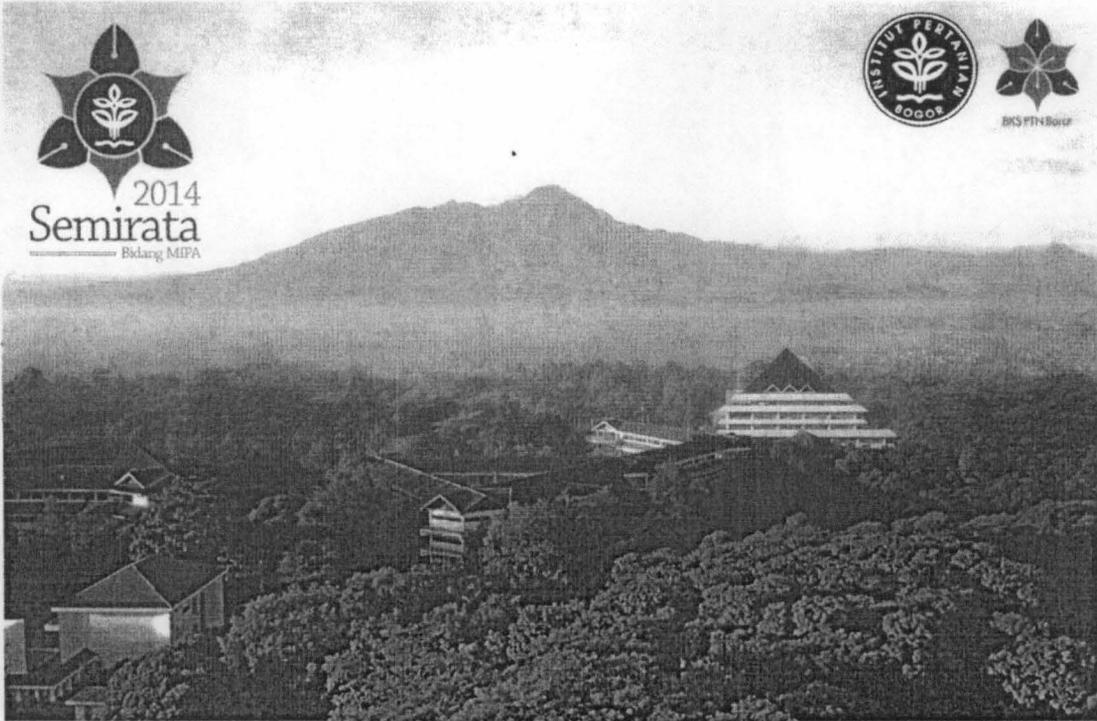


14



PROSIDING

SEMIRATA 2014 Bidang MIPA BKS-PTN-Barat

"Integrasi sains MIPA untuk mengatasi masalah pangan,
energi, kesehatan, reklamasi, dan lingkungan"

IPB International Convention Center dan Kampus IPB Baranangsiang, 9-11 Mei 2014

BUKU 6

**MATEMATIKA, FISIKA, KIMIA, BIOLOGI,
STATISTIKA, KOMPUTER, STEM,
GEOFISIKA DAN METEOROLOGI**

Diterbitkan oleh: Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Pertanian Bogor



ISBN 978-602-70491-0-9

HIERARCHICAL CLUSTERING PADA DATA TIME SERIES HOTSPOT PROVINSI RIAU

HIERARCHICAL CLUSTERING ON HOTSPOT TIME SERIES DATA OF RIAU PROVINCE

Ilham Alpha Dinov^{1*}, Imas Sukaesih Sitanggang²

Departemen Ilmu Komputer, FMIPA, Institut Pertanian Bogor, Bogor^{1*}; ilhamalphadinov@gmail.com;
Kampus IPB Dramaga, Jl. Meranti Wing 20 Level V, Bogor, Indonesia 16680
Departemen Ilmu Komputer, FMIPA, Institut Pertanian Bogor, Bogor²

ABSTRACT

Land and forest fires become yearly issues in Indonesia, particularly in Riau province. This work applied the hierarchical clustering method for hotspot time series data using average distance formula to analyze hotspots distribution in Riau province. This work focuses on the hotspots distribution in June, July and August for the period 2001-2012. Time series data decomposition was generated to obtain the hotspot occurrence patterns over the period. Hierarchical clustering on the data results 4 clusters at the height of dendrogram about 300. Most of hotspots in these clusters are located in plantation.

Keywords: time series data, forest fires, hotspot, hierarchical clustering

ABSTRAK

Kebakaran hutan menjadi permasalahan tahunan di Indonesia khususnya di provinsi Riau. Penelitian ini menerapkan metode *hierarchical clustering* untuk data *time series hotspot* menggunakan rumus jarak rata-rata untuk menganalisis distribusi *hotspot* di provinsi Riau. Penelitian ini fokus pada persebaran *hotspot* di bulan Juni, Juli dan Agustus periode tahun 2001 hingga tahun 2012. Dekomposisi data *time series* diolah untuk mendapatkan pola kemunculan *hotspot* selama periode tersebut. *Hierarchical clustering* menghasilkan 4 *cluster* pada ketinggian dendrogram sekitar 300. Sebagian besar *hotspot* pada *cluster-cluster* tersebut berada di daerah perkebunan.

Katakunci: data time series, hotspot, hierarchical clustering, kebakaran hutan

PENDAHULUAN

Latar Belakang

Kebakaran hutan dan lahan menjadi permasalahan yang serius di Indonesia. Salah satu upaya pencegahan kebakaran hutan tersebut yakni dengan mengetahui persebaran pengelompokan *hotspot* yang berpotensi tinggi terhadap terjadinya kebakaran hutan. Teknik *data mining* sesuai untuk diterapkan pada data *hotspot* tersebut. *Clustering* digunakan untuk untuk mengelompokkan data berdasarkan pada kedekatan dari suatu karakteristik sampel.

Secara umum, *clustering* merupakan proses pengelompokan sekumpulan objek ke dalam kelas yang objeknya mirip [1]. Ukuran kemiripan dan ketidakmiripan dinilai berdasarkan nilai atribut yang mendeskripsikan objek. Tujuan dari *clustering* adalah untuk mengidentifikasi struktur data yang belum berlabel yang ditentukan secara objektif oleh data yang terorganisir ke dalam grup yang homogen dengan meminimalisir kesamaan antar grup dan memaksimalkan ketidakmiripan antar grup [2].

Data *hotspot* merupakan salah satu indikator kemungkinan terjadinya kebakaran hutan pada wilayah tertentu. Pemantauan *hotspot* dilakukan dengan penginderaan jauh

(*remote sensing*) menggunakan satelit. Data *hotspot* merupakan data *time series* karena diamati secara periodik setiap harinya oleh satelit seperti NOAA dan MODIS. *Time series* adalah serangkaian waktu koleksi pengamatan yang dilakukan secara kronologis [3]. Contoh data *time series* yang sering ditemui adalah data suhu harian (*daily temperature*), total penjualan mingguan (*weekly sales totals*) dan harga reksa dana (*prices of mutual funds*).

Penelitian ini dilakukan untuk melihat pola persebaran *hotspot* di Riau dalam kurun waktu dari tahun 2001 sampai tahun 2012. *Time series clustering* dipilih sebagai metode yang digunakan karena merupakan salah satu metode yang sesuai untuk melihat pola pengelompokan per satuan waktu.

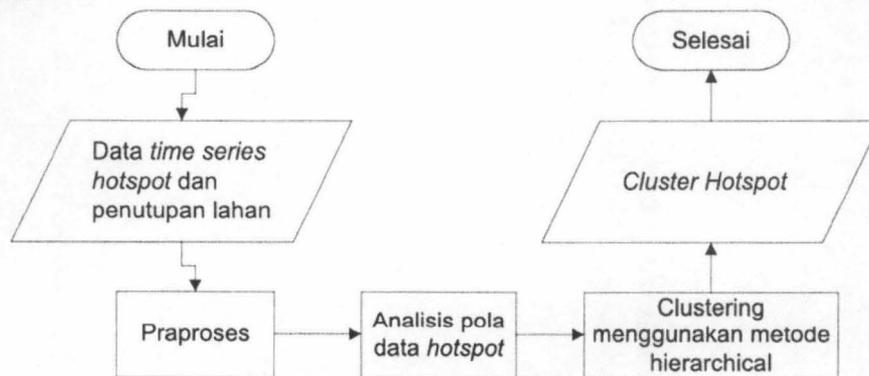
Tujuan, Manfaat dan Ruang Lingkup

Tujuan dari penelitian ini adalah menganalisis data *time series hotspot* dengan melihat pola kecenderungan kemunculan *hotspot* selama 12 tahun dari tahun 2001 hingga tahun 2012, menentukan pengelompokan *hotspot* menggunakan metode *hierarchical clustering*. Penelitian ini diharapkan dapat membantu pihak-pihak yang membutuhkan informasi terkait pengelompokan *hotspot* provinsi Riau periode 2001 – 2012.

Ruang Lingkup dari penelitian ini adalah analisis dan pengelompokan data *time series hotspot* menggunakan modul-modul terkait dalam perangkat lunak komputasi statistika R (www.r-project.org) dan formula jarak yang digunakan adalah rata-rata.

METODE PENELITIAN

Pemodelan dan analisis clustering data *hotspot* dilakukan dengan menggunakan perangkat lunak R. Penelitian ini dilakukan dalam beberapa tahap seperti yang ditunjukkan pada Gambar 1.



Gambar 2 Diagram tahapan penelitian

Data Penelitian

Data yang digunakan dalam penelitian ini adalah data *hotspot* provinsi Riau tahun 2001 hingga tahun 2012 yang diunduh dari situs NASA dan tipe data penutupan lahan provinsi Riau pada tahun 2008 sebagai kelas target. Luas daerah tiap tipe penutupan lahan dapat dilihat pada Tabel 1.

Tabel 2 Kelas penutupan lahan provinsi Riau

Tipe Penutupan Lahan	Luas Area (km ²)
Plantation	28,417.13
Dryland Forest	17,497.66
Agricultural Field	20,294.21
Shrubs	8,176.02
Natural Forest	8,110.41

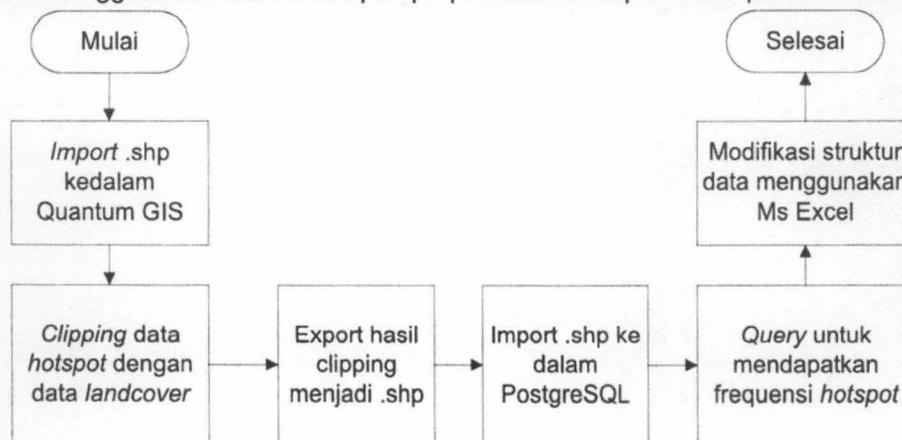
Tipe Penutupan Lahan	Luas Area (km ²)
<i>Bare Land</i>	3,029.42
<i>Swamp & Mangrove</i>	3,746.12

Lingkungan Pengembangan

Penelitian ini dilakukan menggunakan perangkat keras berupa komputer personal dengan spesifikasi *Processor Intel® Core™ i5*, *Memory 4 GB* dan *Hardisk 320 GB*. Sedangkan perangkat lunak yang digunakan adalah Sistem Operasi Microsoft Windows 8, *Tool* untuk komputasi statistika R 3.0.1 64 Bit, *Tool* R Studio untuk memudahkan modifikasi sintaks R, DBMS PostgreSQL 9.1, Quantum GIS 1.8.0 untuk analisis, visualisasi data spasial dan pemisah data *hotspot* per tipe kelas penutupan lahan, PostGIS 2.0 sebagai ekstensi spasial dalam PostgreSQL 9.1 untuk menghubungkan dengan Quantum GIS 1.8.0. dan Microsoft Excel untuk mengolah data *time series* hasil kueri DBMS menjadi masukan R.

Praproses Data

Praproses dilakukan untuk menyiapkan dataset dan mengatur struktur data *time series hotspots* serta data penutupan lahan agar bisa dilakukan kueri pada DBMS PostgreSQL dan diolah menggunakan tool R. Tahapan praproses data dapat dilihat pada Gambar 2.



Gambar 3 Diagram alur praproses data

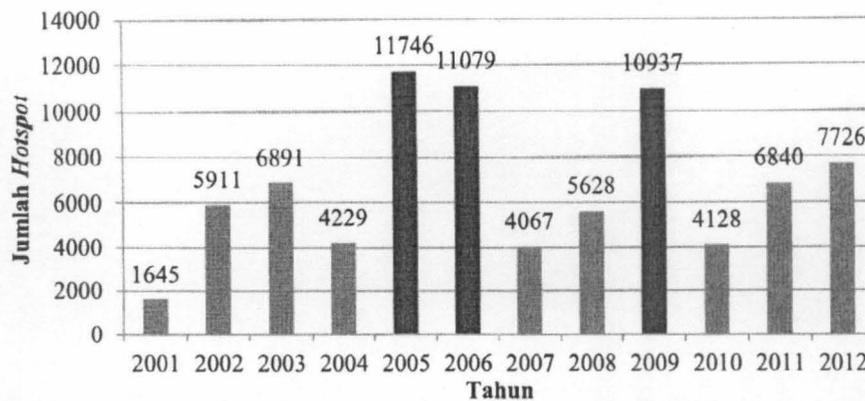
HASIL PEMBAHASAN

Pengumpulan Data

Data persebaran *hotspot*

Data persebaran *hotspot* yang digunakan adalah data *hotspot* dari tahun 2001 hingga tahun 2012 yang diunduh dari situs resmi NASA. Data ini memiliki atribut latitude, longitude dan *acq_date*. Atribut latitude dan longitude menggambarkan letak *hotspot* dalam peta geografis provinsi Riau sedangkan atribut *acq_date* adalah keterangan tanggal suatu *hotspot* muncul pada titik tertentu. Gambar 3 menyajikan jumlah *hotspot* di wilayah provinsi Riau dari tahun 2001 hingga tahun 2012.

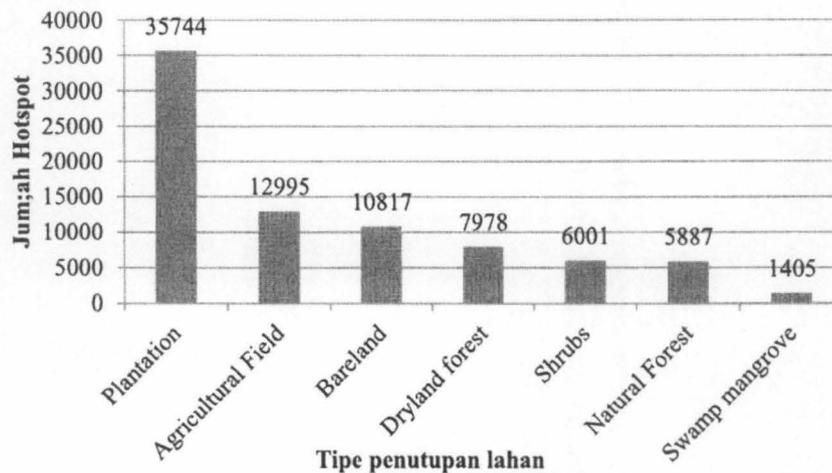
Dari Gambar 3 terlihat bahwa tahun 2005, 2006 dan 2009 merupakan tahun-tahun yang memiliki catatan jumlah *hotspot* tinggi.



Gambar 4 Jumlah *hotspot* provinsi Riau dari tahun 2001 hingga tahun 2012

Data Penutupan Lahan

Data penutupan lahan di wilayah Riau terbagi atas tujuh tipe yaitu lahan perkebunan (*plantation*), hutan kering (*dryland forest*), lahan pertanian (*agricultural field*), semak (*shrubs*), hutan alami (*natural forest*), lahan kosong (*bare land*), serta rawa dan bakau (*swamp & mangrove*). Gambar 4 menunjukkan grafik kemunculan *hotspot* per kelas penutupan lahan dari tahun 2001 hingga tahun 2012.



Gambar 5 Frekuensi *hotspot* berdasarkan tipe penutupan lahan di wilayah Riau tahun 2001 hingga tahun 2012

Praproses Data *Time Series* Hotspot

Import shapefile ke dalam Quantum GIS

Untuk memudahkan memanipulasi data awal yang berbentuk shape file, maka data tersebut di-*import* kedalam *tool* Quantum GIS.

Kueridan *clipping* data *hotspot* dengan data penutupan lahan

Proses *clipping* menggunakan Quantum GIS guna mendapatkan *hotspot* yang berada di wilayah Riau saja. Kemudian dilakukan proses kueri menggunakan PostgreSQL untuk membagi data *hotspot* tiap tahun dari 2001 hingga 2012. Lalu data per tahun tersebut akan dibagi berdasarkan kelas penutupan lahan menggunakan Quantum GIS.

Export layer hasil *clipping* menjadi shape file

Hasil kueri dan *clipping* di-export menjadi *shape file*. Sehingga dapat diolah selanjutnya menggunakan DBMS PostgreSQL.

Import shape file ke dalam PostgreSQL

Terlebih dahulu dipasang *tool* pendukung PostGIS agar PostgreSQL dapat membaca *shape file*. Pada tahap ini akan dihasilkan tabel-tabel *hotspot* provinsi Riau dari tahun 2001 hingga 2012 dan tabel penutupan lahan provinsi Riau.

Kueri untuk mendapatkan frekuensi *hotspot* per hari

Pada proses ini akan dilakukan kueri untuk mendapatkan frekuensi *hotspot* per hari dan per kelas penutupan lahan dalam periode 12 (dua belas) tahun.

Modifikasi struktur data menggunakan Microsoft Excel

Perlu dilakukan modifikasi struktur data *time series hotspot* sehingga dapat menjadi data inputan bagi tool R. Proses ini menggunakan bantuan Microsoft Excel.

Analisis Data

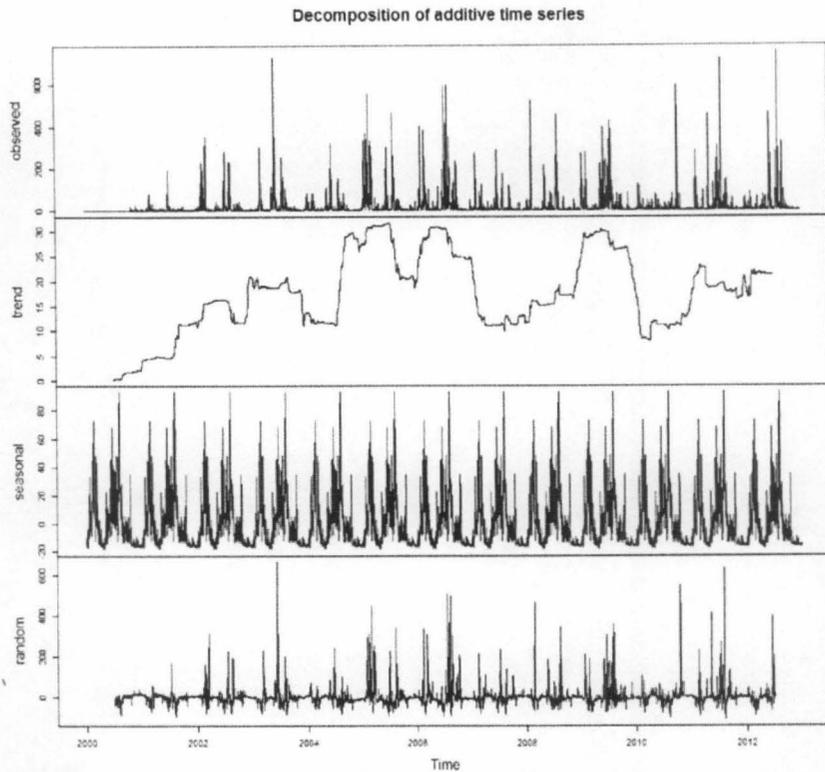
Gambar 5 menyajikan dekomposisi data *time series hotspot* di wilayah Riau pada kurun waktu 2001 hingga 2012. Label *observed* merupakan gambaran banyaknya *hotspot* per hari selama setahun pada data *time series* tahun 2001 hingga tahun 2012. Proses dekomposisi menghasilkan *trend* yang tidak linear maupun kuadratik, terlihat pada Gambar 5 baris kedua label *Trend*. Apabila dilihat *trend* masing-masing tahun, jumlah *hotspot* akan terlihat signifikan naik dan signifikan turun disekitar pergantian tahun. Hal ini dapat dilihat pada awal pergantian tahun 2002, 2003, 2006, 2008, 2009, dan 2011 yang naik secara signifikan. Sedangkan pada awal pergantian tahun 2004, 2005, 2007 dan 2010 yang turun secara signifikan. *Season* atau musiman dari data *time series* didefinisikan sebagai pola yang berulang selama interval waktu yang tetap.

Clustering Data Time Series

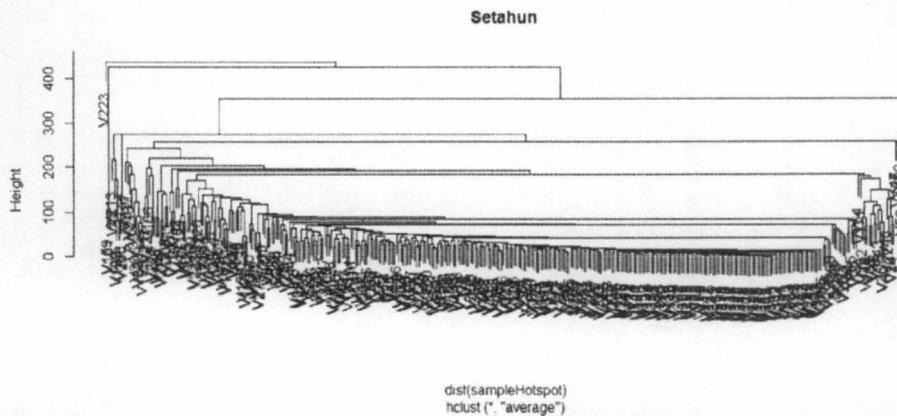
Dalam *clustering*, data akan dikelompokkan menjadi *cluster-cluster* berdasarkan kemiripan satu data dengan yang lain. Prinsip dari *clustering* adalah memaksimalkan kesamaan antar anggota dalam satu *cluster* dan meminimumkan kesamaan antar anggota pada *cluster* yang berbeda.

Dengan menggunakan algoritme *hierarchical clustering* kita bisa membagi *cluster* berdasarkan ketinggian ketidakmiripan atau *dissimilarity* antar *cluster* dendrogram. Hasil keseluruhan dari algoritme *hierarchical clustering* secara grafik dapat digambarkan sebagai *tree*, yang disebut dengan dendrogram. *Tree* ini secara grafik menggambarkan proses penggabungan dari *cluster-cluster* yang ada, sehingga menghasilkan *cluster* dengan level yang lebih tinggi [4]. Tampilan dendrogram hasil *hierarchical clustering* data *hotspot* dalam setahun kurun waktu 2001 hingga tahun 2012 dapat dilihat pada Gambar 6.

Sebanyak 80,852 *hotspot* yang diamati dalam 365 hari yang berada dalam rentang waktu tahun 2001 hingga tahun 2012. Data tersebut dibagi per kelas penutupan lahan tiap tahunnya. Tabel 2 menyajikan rataan dan persentase jumlah *hotspot* tiap kelas penutupan lahan dalam 12 (dua belas) bulan setiap tahun selama kurun waktu dari tahun 2001 hingga tahun 2012.



Gambar 6 Dekomposisi data time series hotspot Riau tahun 2001 hingga tahun 2012

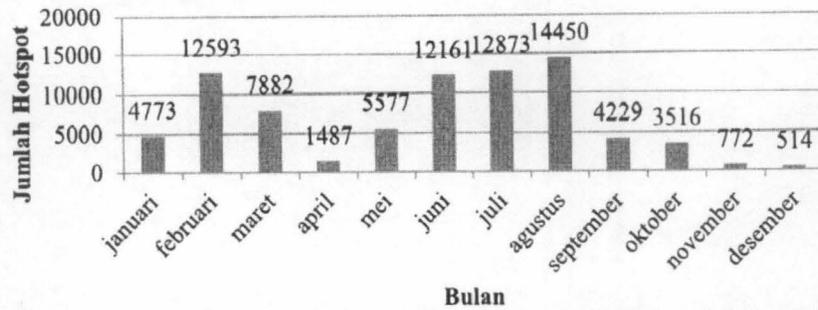


Gambar 7 Dendrogram data hotspot selama setahun dari tahun 2001 hingga tahun 2012
 Tabel 3 Rataan dan Persentase tiap kelas penutupan lahan kurun waktu 2001-2012.

No	Kelas	Rataan hotspot	Persentase (%)
1	Plantation	8.16	44.22
2	Dryland_forest	1.82	9.87
3	Agricultural_field	2.77	16.22
4	Shrubs	1.47	7.28
5	Natural_forest	1.35	7.28
6	Bare_land	2.47	13.38
7	Swamp_Mangrove	0.32	1.74

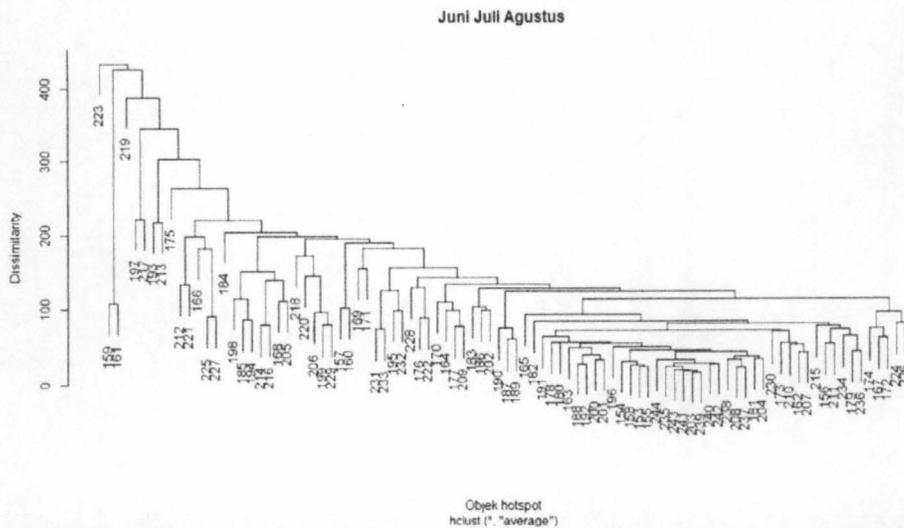
Gambar 7 menyajikan grafik jumlah hotspot tiap bulan selama kurun waktu 12 (dua belas) tahun dari tahun 2001 hingga 2012. Dapat dilihat bahwa bulan Agustus memiliki

frekuensi *hotspot* paling banyak dibandingkan dengan frekuensi *hotspot* di bulan yang lainnya yaitu sebanyak 14.450 *hotspot*.



Gambar 8 Grafik jumlah *hotspot* tiap bulan selama 12 tahun (2001-2012)

Data bulan Agustus dan dua bulan sebelumnya menjadi titik fokus pengamatan karena bulan-bulan tersebut merupakan urutan bulan yang memiliki jumlah frekuensi *hotspot* paling banyak dibandingkan dengan total frekuensi pada urutan bulan-bulan lainnya. Sebanyak 39,484 *hotspot* ditemukan pada bulan-bulan tersebut. Gambar 8 menyajikan tampilan plot pada bulan Juni, Juli dan Agustus selama kurun waktu 12 (dua belas) tahun dari tahun 2001 hingga tahun 2012.



Gambar 9 Dendrogram *hotspot* bulan Juni, Juli dan Agustus dari tahun 2001 hingga 2012

Dari Gambar 8 dapat dilihat bahwa proses *clustering* menghasilkan dendrogram yang memiliki ketinggian yang menyatakan ketidaksamaan atau *dissimilarity* antar dua *cluster*. Pada Tabel 3 dapat terlihat pembagian *cluster* berdasarkan ketinggian dendrogram dan susunan objek anggota tiap-tiap *cluster* tersebut.

Tabel 4 Cluster berdasarkan ketinggian dendrogram

Ketinggian Dendrogram	Banyak Cluster	Anggota	
300	4	Cluster 1	159, 161
		Cluster 2	197, 217
		Cluster 3	193, 213
		Cluster 4	175, 212, 221, 166, 225, 227, 184, 198, 185, 194, 214, 216, 168, 205, 218, 220, 206, 199, 229, 157, 160, 169, 171, 231, 233, 195, 232, 228, 176, 222, 170, 164, 177, 209, 183, 186, 202, 190, 187, 189, 165, 182, 191, 178, 180, 163, 188, 192, 200, 201, 196, 154, 158, 153, 155, 244, 235, 243, 241, 203, 239, 240, 242, 238, 208, 237, 181, 204, 230, 173, 210, 162, 207, 215, 156, 211, 234, 179, 236, 174, 167, 172, 224, 226
200	4	Cluster 1	159, 161
		Cluster 2	212, 221, 166, 225, 227
		Cluster 3	198, 185, 194, 214, 216, 168, 205
		Cluster 4	218, 220, 206, 199, 229, 157, 160, 169, 171, 231, 233, 195, 232, 228, 176, 222, 170, 164, 177, 209, 183, 186, 202, 190, 187, 189, 165, 182, 191, 178, 180, 163, 188, 192, 200, 201, 196, 154, 158, 153, 155, 244, 235, 243, 241, 203, 239, 240, 242, 238, 208, 237, 181, 204, 230, 173, 210, 162, 207, 215, 156, 211, 234, 179, 236, 174, 167, 172, 224, 226

Pada Tabel 3 dapat dilihat pemotongan dendrogram pada titik ketinggian 300 menghasilkan sebanyak 4 cluster dan pemotongan dendrogram pada titik ketinggian 200 juga menghasilkan 4 cluster.

Tabel 4 menyajikan persebaran jumlah hotspot per cluster pada pemotongan titik 300 dendrogram pada data bulan Juni, Juli dan Agustus periode tahun 2001 hingga 2012 untuk setiap tipe penutupan lahan. Tipe penutupan lahan *plantation* (perkebunan) memiliki frekuensi tertinggi pada tiap cluster yang telah dihasilkan.

Tabel 5 Jumlah hotspot tiap cluster dengan pemotongan dendrogram pada ketinggian 300 per tipe penutupan lahan

Tipe Penutupan Lahan	Jumlah Hotspot			
	Cluster 1	Cluster 2	Cluster 3	Cluster 4
<i>Plantation</i>	279	604	557	16,476
<i>Dryland Forest</i>	80	146	160	3,757
<i>Agricultural Field</i>	95	254	212	6,011
<i>Shrubs</i>	61	157	82	3,233
<i>Natural Forest</i>	34	130	179	3,368
<i>Bareland</i>	42	134	55	2,748
<i>Swamp mangrove</i>	11	33	17	569

Tabel 5 menyajikan persebaran jumlah hotspot pada masing-masing cluster per tipe penutupan lahan pada pemotongan titik ketinggian 300 pada dendrogram data bulan Juni, Juli, dan Agustus periode tahun 2001 hingga 2012. Tipe penutupan lahan *plantation* (perkebunan) memiliki frekuensi tertinggi pada tiap cluster yang telah dihasilkan.

Tabel 6 Jumlah hotspot tiap cluster dengan pemotongan dendrogram pada ketinggian 200 per tipe penutupan lahan

Tipe penutupan lahan	Jumlah Hotspot			
	Cluster 1	Cluster 2	Cluster 3	Cluster 4
<i>Plantation</i>	279	1,266	1,629	14,742
<i>Dryland forest</i>	80	383	269	3,411
<i>Agricultural Field</i>	95	527	715	5,235
<i>Shrubs</i>	61	214	408	2,850
<i>Natural Forest</i>	34	224	318	3,135
<i>Bareland</i>	42	172	295	2,470
<i>Swamp mangrove</i>	11	55	51	513

SIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan terhadap data *time series hotspot*, diambil beberapa kesimpulan. Pertama, jumlah *hotspot* signifikan naik dan signifikan turun di sekitar pergantian tahun. Kedua, pada pemotongan titik ketinggian dendrogram 300 dari data bulan Juni, Juli dan Agustus hasilkan 4 buah *cluster* dan 4 buah *cluster* pada pemotongan di titik ketinggian dendrogram 200. Ketiga, pada pemotongan ketinggian dendrogram 300 dari data bulan Juni, Juli dan Agustus tipe penutupan lahan perkebunan (*plantation*) memiliki jumlah *hotspot* paling banyak di setiap *cluster*. Keempat, pada pemotongan ketinggian dendrogram 200 untuk data bulan Juni, Juli dan Agustus tipe penutupan lahan perkebunan (*plantation*) memiliki jumlah *hotspot* paling banyak di setiap *cluster*. Kelima, dalam kurun waktu 12 tahun persentase kemunculan *hotspot* terbanyak pada penutupan lahan perkebunan (*plantation*).

Beberapa hal yang perlu dikembangkan lebih lanjut dari penelitian ini yaitu penggunaan metode *clustering* lain guna mendapatkan hasil *cluster* yang lebih baik dan dilakukan validasi *cluster* untuk mendapatkan *cluster* terbaik dari data *time series hotspot*.

DAFTAR PUSTAKA

- [1] Han J, Kamber M. 2006. *Data Mining Concepts and Techniques Second Edition*. San Fransisco (US): Morgan Kaufmann Publisher.
- [2] Liao TW. 2005. Clustering of time series data—a survey. *Pattern Recognition*. 38(11): 1857-1874.
- [3] Tak-Chung Fu. 2011. A review on time series data mining. *Engineering Applications of Artificial Intelligence*. 24 (1): 164–181.
- [4] Steinbach M, Karypis G, Kumar V. 2000. *A comparison of document clustering techniques*. Minnesota (US) [Internet]. [diunduh 2014 Jan 01]. Tersedia pada: <http://glaros.dtc.umn.edu/gkhome/fetch/papers/doccluster.pdf>