



ANALISIS DATA PUBLIKASI DENGAN PENDEKATAN BIBLIOMETRIKS DAN TEXT CLUSTERING UNTUK MENENTUKAN KRITERIA TINGKAT KEPARAHAAN KARHUTLA

MUHAMMAD IQBAL ALVIAN WIDHYARTO



**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



IPB University

@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

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

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

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





PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Analisis Data Publikasi Dengan Pendekatan Bibliometriks dan *Text Clustering* Untuk Menentukan Kriteria Tingkat Keparahan Karhutla” adalah karya saya dengan arahan dari dosen 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, Juli 2024

Muhammad Iqbal Alvian Widiyarto
G6401201066



ABSTRAK

MUHAMMAD IQBAL ALVIAN WIDIYARTO. Analisis Data Publikasi dengan Pendekatan Bibliometriks dan *Text Clustering* Untuk Menentukan Tingkat Keparahan Karhutla. Dibimbing oleh HARI AGUNG ADRIANTO dan IMAS SUKAESIH SITANGGANG.

Kebakaran hutan dan lahan (karhutla) di Indonesia telah menjadi isu lingkungan selama dua dekade terakhir, dengan dampaknya yang melibatkan berbagai aspek. Penilaian dampak kebakaran terutama dilakukan untuk penelitian, perencanaan rehabilitasi hutan, konservasi, dan penegakan hukum. Namun, standar penilaian dampak kebakaran yang digunakan sebagai acuan pengendalian karhutla masih sangat beragam. Penelitian ini bertujuan untuk mengidentifikasi kriteria untuk mengukur tingkat keparahan kebakaran hutan dengan menerapkan algoritma Bibliometriks dan *text mining*. Selain itu, analisis bibliometriks akan dibandingkan dengan hasil dari analisis *clustering*. Penelitian ini membandingkan beberapa algoritma *clustering* seperti *Regular K-Means*, *Bisecting K-Means*, dan *unweighted pair group method with arithmetic mean* (UPGMA) *Clustering* untuk menganalisis data artikel literatur dari tahun 2013 hingga 2023. Setiap algoritma dievaluasi menggunakan *silhouette score* dan jumlah *cluster* yang berbeda. *Regular K-Means* menghasilkan 10 *cluster* dengan *silhouette score* 0,138. *Bisecting K-Means* menghasilkan *silhouette score* 0,067 dengan jumlah *cluster* sebanyak 9. UPGMA *Clustering* menghasilkan *silhouette score* 0,138 dengan 10 *cluster*. Analisis bibliometriks menunjukkan *burned area* dan *peatland* menjadi kriteria yang muncul pada hasil bibliometriks dengan masing-masing berada di kuadran yang beririsan antara 1 dan 4, dan kuadran 1 yang memiliki arti bahwa kriteria terkait merupakan tema yang penting untuk penelitian dikarenakan banyaknya jumlah penelitian terkait dan terdapat hubungan yang besar dengan penelitian lainnya dan kuadran 4 memiliki arti bahwa kriteria terkait merupakan kriteria yang tidak banyak diteliti akan tetapi banyak hubungannya dengan penelitian lainnya. Analisis hasil *clustering* menunjukkan kriteria dan indikator berupa *burned area*, *peatland*, dan *vegetation* sebagai kriteria dan indikator terbanyak untuk hasil dari metode Bibliometriks dan *clustering*.

Kata Kunci: Algoritma K-Means, *Clustering*, Bibliometriks, Kebakaran hutan dan lahan, *Text mining*

ABSTRACT

MUHAMMAD IQBAL ALVIAN WIDIYARTO. Publication data analyst with Bibliometrics and Text Clustering Approach to Identify the Criteria of Forest and Land Fire Severity. Supervised by HARI AGUNG ADRIANTO and IMAS SUKAESIH SITANGGANG.

Forest and land fires in Indonesia have been an environmental issue for the past two decades, with their impacts involving various aspects. Fire impact assessments are mainly conducted for research, forest rehabilitation planning, conservation and law enforcement. However, fire impact assessment standards used



as a reference for forest and land fire control are still very diverse. This research aims to identify criteria for measuring the severity of forest fires by applying Bibliometrics and text mining algorithms. In addition, the bibliometric analysis will be compared with the results of clustering analysis. This research compares several clustering algorithms such as Regular K-Means, Bisecting K-Means, and unweighted pair group method with arithmetic mean (UPGMA) Clustering to analyze literature article data from 2013 to 2023. Each algorithm was evaluated using different silhouette scores and number of clusters. Regular K-Means resulted in 10 clusters with a silhouette score of 0.138. Bisecting K-Means produced a silhouette score of 0.067 with a cluster count of 9. UPGMA Clustering produced a silhouette score of 0.138 with 10 clusters. Bibliometric analysis shows that burned area and peatland are the criteria that appear in the bibliometric results with each being in quadrants that intersect between 1 and 4, and quadrant 1 which means that related criteria are important themes for research due to the large number of related studies and there is a large relationship with other studies and quadrant 4 means that related criteria are criteria that are not widely researched but have a lot to do with other studies. Analysis of the clustering results shows criteria and indicators such as burned area, peatland, and vegetation as the most criteria and indicators for the results of the Bibliometrics and clustering methods.

Keywords: Clustering, Forest and Land Fire, Bibliometrics, K-Means Algorithm, Text Mining



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

@Hak cipta milik IPB University

IPB University

© Hak Cipta milik IPB, tahun 2024
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

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



**ANALISIS DATA PUBLIKASI DENGAN PENDEKATAN
BIBLIOMETRIKS DAN TEXT CLUSTERING UNTUK
MENENTUKAN KRITERIA TINGKAT KEPARAHAN KARHUTLA**

MUHAMMAD IQBAL ALVIAN WIDHYARTO

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Ilmu komputer

**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



Penguji pada ujian Skripsi:

Muhammad Asyhar Agmalaro, S.Si, M.Kom

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa menantumkan 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 University.

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



Judul : Analisis Data Publikasi Dengan Pendekatan Bibliometriks dan *Text Clustering* Untuk Mengukur Kriteria Tingkat Keparahan Karhutla
Nama : Muhammad Iqbal Alvian Widiyarto
NIM : G6401201066

Disetujui oleh



Pembimbing 1:
Hari Agung Adrianto, S.Kom, M.Si., Ph.D.

Pembimbing 2:
Prof. Dr. Imas Sukaesih Sitanggang, S.Si., M.Kom.



Diketahui oleh

Ketua Departemen Ilmu Komputer:
Dr. Sony Hartono Wijaya, S.Kom., M.Kom.
19810809 200812 1 002



Tanggal Ujian:
23 Juli 2024

Tanggal Lulus:



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

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah SWT atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan November 2023 sampai bulan Juli 2024 ini ialah kebakaran hutan dan lahan (karhutla), dengan judul “Analisis Data Publikasi Dengan Pendekatan Bibliometriks dan *Text Clustering* Untuk Mengukur Kriteria Tingkat Keparahan Karhutla”. Terima kasih penulis ucapan kepada orang tua tercinta yaitu Bapak Helmi Nur Widiyarto dan Ibu Neni Sofiani serta dosen pembimbing Prof. Dr. Imas Sukaesih Sitanggang, S.Si., M.Kom dan Hari Agung Adrianto, S.Kom., M.Si., Ph.D yang telah membimbing dan banyak memberi saran. Ucapan terima kasih juga disampaikan kepada pembimbing akademik, moderator seminar, dan penguji luar komisi pembimbing.

Ungkapan terima kasih penulis juga disampaikan kepada teman-teman Kom ST10, Dcampus, Etherone 57, rekan-rekan AURI dan FPI, beserta orang-orang baik yang sudah menemani dan bersama-sama selama kuliah di IPB. Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan

DAFTAR ISI

ABSTRAK	ii
DAFTAR TABEL	vii
DAFTAR GAMBAR	vii
I. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Ruang Lingkup	4
II. TINJAUAN PUSTAKA	4
2.1 Kebakaran Hutan dan Lahan	4
2.2 <i>Literature Review</i>	4
2.3 <i>Text Mining</i>	6
2.4 <i>Clustering</i>	7
2.5 Bibliometriks	9
2.6 <i>Linkage</i>	10
III. METODE	13
3.1 Data Penelitian	13
3.2 Tahapan Penelitian	15
3.3 Lingkungan Pengembangan	16
IV. HASIL DAN PEMBAHASAN	16
4.1 Menggumpulkan Artikel Literatur	16
4.2 Ekstraksi Data	17
4.3 Analisis Bibliometriks	18
4.4 <i>Clustering</i>	20
4.5 Evaluasi dan Komparasi Metode	25
4.5 Analisis Kriteria dan Indikator Tingkat Keparahan Karhutla	30
V. SIMPULAN DAN SARAN	33
5.1 Simpulan	33
5.2 Saran	33
DAFTAR PUSTAKA	35
LAMPIRAN	38



DAFTAR TABEL

1	<i>Tools</i> pendukung <i>literature review</i>	5
2	Term dan frekuensi 10 teratas	17
3	Frekuensi tema penelitian teratas	19
4	Komparasi formula <i>linkage</i> algoritma UPGMA	23
5	Karakteristik dari <i>cluster</i> algoritma <i>regular K-Means</i>	25
6	Karakteristik dari <i>cluster</i> algoritma <i>bisecting K-Means</i>	27
7	Karakteristik dari <i>cluster</i> algoritma UPGMA <i>clustering</i>	29
8	Analisis kriteria tingkat keparahan Karhutla	31

DAFTAR GAMBAR

1	Algoritma <i>Regular K-Means</i> (Banerjee <i>et al.</i> 2015)	7
2	Algoritma <i>bisecting K-Means</i> (Banerjee <i>et al.</i> 2015)	8
3	Algoritma UPGMA (Yoshida <i>et al.</i> 2022)	9
4	Kuadran thematic map dalam analisis bibliometriks (Yu dan Munoz 2020)	10
5	Diagram alur penelitian	13
6	Frekuensi artikel per tahun yang digunakan	16
7	Penelusuran dengan Publish or Perish	17
8	<i>Wordcloud</i> kumpulan artikel literatur	18
9	Hasil <i>text extraction</i>	18
10	Hasil <i>thematic map network</i> dengan menggunakan Biblioshiny	19
11	pembagian kuadran hasil analisis bibliometriks untuk setiap <i>cluster</i>	20
12	<i>Silhouette score</i> dari algoritma <i>regular K-Means</i>	21
13	Sebaran <i>cluster</i> dari algoritma <i>regular K-Means</i>	21
14	<i>Silhouette score</i> dari algoritma <i>Bisecting K-Means</i>	22
15	Sebaran <i>cluster</i> dari algoritma <i>Bisecting K-Means</i>	23
16	<i>silhouette score</i> dari algoritma UPGMA <i>Clustering</i>	24
17	Sebaran <i>cluster</i> dari algoritma UPGMA <i>clustering</i>	24

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa memantulkan 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 University.

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



1.1 Latar Belakang

Indonesia adalah sebuah negara kepulauan dengan iklim tropis yang memiliki 17.000 pulau dengan luas daratan yang mencapai 1,9 juta km² (Fadilah dan Adhi 2022). Menurut data dari Kementerian Lingkungan Hidup dan Kehutanan (KLHK), pada tahun 2022, Indonesia memiliki kawasan hutan seluas 125,76 hektar. Hutan merupakan beraneka ragam jenis tumbuhan dan hewan yang bersekutu dalam asosiasi biotis dimana asosiasi tersebut bersama lingkungan di sekitarnya membentuk sistem ekologis dan organisme yang saling mempengaruhi sehingga terbentuk siklus energi yang kompleks (Spurr dan Burton 1973). Berdasarkan fungsinya, kawasan hutan lindung menjadi yang paling luas di Indonesia, yakni 29,56 juta ha. Luasan tersebut setara dengan 23,5% dari total kawasan hutan secara nasional. Kemudian, luas kawasan hutan produksi tetap sebesar 29,23 juta ha. Lalu, kawasan hutan yang masuk ke dalam konservasi memiliki luas 27,41 juta ha. Luas hutan produksi terbatas sebesar 26,8 juta ha. Sedangkan, hutan produksi yang dapat dikonversi memiliki luas 12,79 juta ha (Widi 2023).

Kebakaran hutan dan lahan atau sering disebut juga dengan karhutla terjadi di banyak tempat di dunia termasuk Indonesia. Karhutla di Indonesia merupakan permasalahan lingkungan hidup yang terjadi setiap tahun selama dua dekade terakhir, dimana dampak dari karhutla di Indonesia tidak hanya dirasakan oleh masyarakat indonesia saja melainkan hingga dirasakan oleh masyarakat negara tetangga. Menurut data Kementerian Lingkungan Hidup dan Kehutanan. Selama periode Januari-Juli 2023 luas karhutla di Indonesia sudah mencapai 642.099 hektar (BPS 2023). Dalam periode Januari-Juli 2023, provinsi penghasil Karhutla terbesar adalah Nusa Tenggara Timur sebesar 28.718 ha, Kalimantan Barat sebesar 12.537 ha, dan Nusa Tenggara Barat sebesar 9.662 ha. Menurut penelitian oleh Loren *et al.* (2015), penyebab kebakaran yang utama ada lima, yang pertama adalah pembukaan lahan baru dengan cara pembakaran lahan yang tak terkendali, kemudian membuang puntung rokok sembarangan, bahan-bahan kering yang menjadi bahan bakar yang dapat mudah terbakar, loncatan panas api yang datang dari daerah dan wilayah lain, dan faktor alam seperti musim kemarau yang panas sehingga terjadi gesekan benda kering yang dapat mudah terbakar.

Berdasarkan peraturan yang berlaku di Indonesia yang bersumber dari MLHK dalam Peraturan P.32/MenLHK/Setjen/Kum.1/3/2016 mengenai Pengendalian Karhutla, penilaian dampak dari kebakaran hutan merupakan salah satu tindakan dalam pengendalian kebakaran hutan. Hal ini mengacu pada Peraturan Pemerintah Nomor 23 Tahun 2021 tentang Penyelenggaraan Kehutanan Pasal 258 ayat 2 tentang kegiatan identifikasi dan evaluasi dari kebakaran hutan. Kegiatan yang dimaksud dalam peraturan ini meliputi pengumpulan data dan informasi mengenai kebakaran, pengukuran dan pembuatan sketsa lokasi kebakaran, dan analisis serta rekomendasi tingkat kerusakan. Umumnya, Penilaian dampak kebakaran dilakukan untuk penelitian dan rehabilitasi hutan, konservasi hutan, dan penegakan hukum. Namun, penilaian dampak kebakaran terstandarisasi yang dijadikan acuan pengendalian kebakaran hutan masih kurang optimal dikarenakan metode yang digunakan dalam penilaian lapangan bervariasi. Salah satu pendekatan dalam menilai dampak kebakaran adalah tingkat keparahan kebakaran dimana pendekatan ini mengukur dampak kebakaran terhadap

ekosistem. Besarnya dampak kebakaran terhadap ekosistem dan lingkungan hidup bergantung pada beberapa faktor, yaitu: kebakaran intensitas, tingkat keparahan kebakaran, jenis tanah, hujan yang turun setelah kebakaran, dan luasnya kebakaran (Syaufina dan Ainuddin 2011).

Perkembangan terkini menunjukkan cara pengukuran dari tingkat keparahan kebakaran yang berbeda-beda. Sebagian besar studi yang mencoba mengukur tingkat keparahan kebakaran mempunyai kesamaan dasar yang berpusat pada hilangnya bahan organik terbaik di atas permukaan tanah maupun di bawah tanah. Metrik di atas permukaan tanah seperti volume tajuk hangus yang digunakan pada hutan atau diameter ranting tersisa di cabang pohon di hutan dan semak belukar, ini dapat disebut juga indikator hilangnya biomassa (Keeley 2009). Dalam Syaufina (2017), tingkat keparahan kebakaran dapat juga dilihat berdasarkan jenis atau kelompoknya, yaitu berdasarkan kondisi tanah dan sifat-sifatnya (*land*) yang dikemukakan oleh Hungerford (1996), persentase total area yang terbakar (*burned area*) yang dikemukakan oleh Wells *et al.* (1979), kerusakan pohon (*vegetation*) yang dikemukakan oleh Foliot dan Bennet (1996), dan lapisan gambut (*peatland*) yang terbakar yang dikemukakan oleh Arstybashev (1983).

Sebelumnya, terdapat penelitian untuk menentukan kriteria dan indikator dari tingkat keparahan kebakaran hutan dengan pola pada sumber teks atau *text mining*. Penelitian sebelumnya yang dilakukan oleh Afina *et al.* (2021) menggunakan VOSviewer untuk mengidentifikasi variabel kriteria dan indikator keparahan kebakaran hutan. Dalam penelitian ini, dari 269 artikel literatur yang diolah, didapatkan hasil berupa 4 *cluster* aspek dari indikator keparahan karhutla dengan kriteria yang didapatkan adalah vegetasi dan tanah dengan indikatornya adalah keberagaman, kelimpahan, struktur hutan, tingkat kematian pohon, dan tingkat kedalaman kebakaran.

Pada penelitian yang dilakukan oleh Afina *et al.* (2021), meskipun mereka menggunakan metode VOSviewer dan jaringan bibliometriks, terdapat kekurangan dalam penjelasan mendalam mengenai penerapan algoritma *text mining* atau *clustering* seperti contohnya K-Means, ECLAT, dan HFTC. Penjelasan lebih lanjut mengenai parameter yang digunakan, langkah-langkah implementasi algoritma, dan bagaimana interpretasi hasil yang didapatkan akan memberikan pemahaman yang lebih mendalam. Penelitian dengan menggunakan *clustering* ini menjadi penting karena diperlukan pengelompokan yang lebih jelas dan terstruktur dalam menganalisis artikel yang dijadikan bahan penelitian. Hal ini diperlukan terutama untuk menentukan indikator mana yang paling signifikan dalam menentukan indikator tingkat keparahan kebakaran hutan.

Text clustering dalam *text mining* dianggap efektif untuk mengelompokkan dan melihat pola kategori yang banyak digunakan dalam pengukuran tingkat keparahan kebakaran hutan. Dengan berbagai cara pengukuran yang berbeda, *text clustering* dapat membantu dalam mengidentifikasi kategori yang paling relevan dan sering digunakan. Menurut Yudiatara *et al.* (2018), *text clustering* bertujuan untuk mengelompokkan data dengan karakteristik yang sama ke dalam kelompok yang sama, dan data dengan karakteristik yang berbeda ke dalam kelompok yang berbeda. Hal ini sangat berguna dalam konteks pengukuran kebakaran hutan di mana data yang ada sangat bervariasi.

Selain itu, Kambey *et al.* (2020) menyebutkan bahwa *text clustering* merupakan tugas penting dalam *text mining* yang melibatkan pengkategorisasian

dan pengelompokan teks. Tujuan utama dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Dengan menggunakan teknik-teknik *text mining*, khususnya *text clustering*, analisis data besar yang berasal dari berbagai sumber bisa dilakukan dengan lebih efektif dan efisien. *Text mining* memungkinkan para peneliti dan praktisi untuk mengekstrak informasi penting yang dapat membantu dalam pengambilan keputusan terkait mitigasi dan penanganan kebakaran hutan.

Penelitian ini menggunakan bahasa pemrograman R untuk melakukan analisis *text mining*. Sumber-sumber artikel literatur yang digunakan juga lebih diperbarui dengan menggunakan artikel-artikel literatur terbaru, sehingga hasil dapat mencerminkan perkembangan terkini dalam penelitian mengenai indikator dan kriteria tingkat keparahan karhutla. Penelitian ini mencoba berbagai algoritma *clustering* yang beragam dan hasil dari masing-masing algoritma dibandingkan untuk menentukan algoritma *clustering* yang memberikan hasil yang relevan dalam konteks penelitian ini. Pendekatan ini memberikan kejelasan tambahan terkait faktor-faktor apa yang paling berpengaruh terhadap keparahan karhutla.

1.2 Rumusan Masalah

Berdasarkan latar belakang, rumusan masalah yang dapat diambil adalah Bagaimana cara mengekstrak data berupa teks yang diambil dari artikel literatur dimana data yang diperoleh akan diolah untuk dilakukan proses *data mining* menggunakan bahasa pemrograman R untuk menentukan apa saja kriteria dan indikator penilaian tingkat keparahan Karhutla.

1.3 Tujuan

Berdasarkan rumusan masalah, maka tujuan dari penelitian ini adalah:

1. Menganalisis jaringan bibliometriks dari hasil publikasi terkait tingkat keparahan area pasca karhutla
2. Menentukan kriteria tingkat keparahan pasca karhutla dengan cara menggunakan algoritma *clustering*
3. Melakukan verifikasi untuk menentukan kriteria tingkat keparahan karhutla bedasarkan bibliometriks dan *clustering*

1.4 Manfaat

Manfaat dari Penelitian ini adalah menghasilkan indikator untuk tingkat keparahan karhutla berdasarkan hasil artikel literatur yang sudah ada. Kemudian Indikator dari penelitian ini diharapkan dapat menjadi salah satu acuan untuk pengukuran tingkat keparahan karhutla kedepannya.

1.5 Ruang Lingkup

Ruang lingkup cakupan untuk penelitian ini dibatasi hanya untuk karhutla yang berada di wilayah Indonesia. Artikel dikumpulkan dari jurnal bereputasi terindeks scopus dengan quartil 1, quartil 2, quartil 3, dan quartil dan Jurnal dengan reputasi Sinta 1, 2, dan 3. Data artikel yang digunakan adalah artikel yang berasal dari literatur jurnal, konferensi, atau publikasi terindeks Scopus. Artikel literatur dipublikasikan dari 2013 hingga 2023. Artikel publikasi yang digunakan dalam penelitian ini bersifat *open access*.

II. TINJAUAN PUSTAKA

2.1 Kebakaran Hutan dan Lahan

Kebakaran hutan dan lahan (karhutla) merupakan peristiwa terbakarnya hutan dan/atau lahan, baik yang terjadi secara alami maupun yang terjadi oleh perbuatan sengaja manusia sehingga mengakibatkan terjadinya kerusakan lingkungan yang dapat menimbulkan kerugian ekologi, ekonomi, sosial, budaya dan politik (Ivandy 2023). Karhutla juga dapat dibedakan pengertiannya dengan kebakaran lahan, dimana perbedaannya antara keduanya terletak pada lokasi kejadiannya. Kebakaran hutan yaitu kebakaran yang terjadi di dalam kawasan hutan, sedangkan kebakaran lahan adalah kebakaran yang terjadi diluar kawasan hutan (Pubowaseso 2004). Keparahan kebakaran (*fire severity*) didefinisikan sebagai istilah untuk menjelaskan respons ekosistem terhadap api dan dapat digunakan untuk menggambarkan pengaruh api terhadap tanah, sistem air, ekosistem, flora dan fauna, atmosfer, serta masyarakat (Syaufina 2017).

Kebakaran hutan dapat dikelompokkan pada tiga tipe, dimana pengelompokan tersebut didasarkan kepada bahan bakar yang mendominasi kebakaran hutan. Tiga tipe bakaran menurut Syaufina (2008), yaitu: kebakaran bawah, kebakaran permukaan, dan kebakaran tajuk. Kebakaran bawah merupakan sebuah situasi dimana api membakar bahan-bahan organik di bawah permukaan serasah, Kebakaran permukaan yaitu situasi dimana api membakar serasah, tumbuhan bawah, bekas limbah pembalakan dan bahan bakar lain yang terdapat di lantai hutan, kebakaran tajuk yaitu situasi dimana api menjalar dari tajuk pohon satu ke tajuk pohon yang lain yang saling berdekatan.

Karhutla dapat terjadi karena pembakaran yang tidak dikendalikan, karena proses spontan alami, atau karena kesengajaan. Proses alami sebagai contohnya kilat yang menyambar pohon atau bangunan, letusan gunung api yang menebarkan bongkahan bara api, dan gesekan antara ranting tumbuhan kering yang mengandung minyak karena goyangan angin yang menimbulkan panas atau percikan api (Notohadinegoro 2006).

2.2 Literature Review

Literature review dapat dideskripsikan sebagai sebuah metode sistematik yang bertujuan untuk mengumpulkan dan memilah-milah suatu penelitian atau teks yang sudah ada sebelumnya (Baumeister dan Leary 1997). Sebuah *Literature review* yang baik dapat dilakukan dengan mengambil dan menyatukan dari berbagai sumber yang didapatkan, dapat menjawab banyak pertanyaan dalam penelitian dan dapat juga memperkokoh suatu teori pengetahuan (Webster dan Watson 2002). *Literature review* membahas mengenai informasi yang dipublikasikan dalam bidang subjek tertentu, dan terkadang informasi dalam bidang studi tertentu dalam jangka waktu tertentu (Amin *et al.* 2014). Tujuan utama *literature review* adalah untuk memberikan informasi kepada pembaca literatur tentang suatu topik dan dapat menjadi sebuah dasar untuk tujuan lain (Cronin *et al.* 2001).

Snyder (2019) menjelaskan bahwa gambaran besar langkah-langkah untuk melakukan suatu *literature review* terdiri 4 tahapan, tahapan yang pertama adalah *design* dimana tahap awal ini adalah untuk mengetahui target dari review, berikutnya adalah tahap *conduct* yaitu sebuah proses *review* dengan menggunakan sumber-sumber yang terpercaya yang telah dicari, yang ketiga adalah tahap *analysis*



untuk mengekstrak atau mengambil informasi yang diperlukan, dan tahapan yang terakhir adalah proses *structuring and writing review* yang merupakan proses integrasi informasi-informasi yang telah didapatkan. Menurut Templier dan Pare (2015), langkah-langkah yang digunakan untuk melakukan *literature review* adalah merumuskan pertanyaan dan tujuan *review*, mencari literatur, penyaringan literatur, menilai kualitas literatur yang didapatkan, ekstraksi data, dan analisis data literatur. *Literature review* juga dapat dilakukan dengan *tools-tools* seperti yang ditunjukkan pada Tabel 1.

Tabel 1 *Tools* pendukung *literature review*

Tools	Deskripsi
Publish or Perish	Publish or Perish adalah program <i>software</i> yang mengambil dan menganalisis artikel akademik. <i>Software</i> ini dapat menggunakan berbagai sumber data untuk memperoleh kutipan mentah, kemudian menganalisisnya dan menyajikan dalam bentuk serangkaian metrik kutipan. Publish or Perish dapat digunakan untuk memilah jurnal, untuk mempersiapkan wawancara kerja, untuk melakukan <i>literature review</i> , untuk melakukan analisis bibliometriks, atau untuk mengerjakan pekerjaan rumah
Biblioshiny	Biblioshiny adalah aplikasi yang menyediakan <i>interface web</i> untuk analisis bibliometriks. Biblioshiny mendukung dalam menggunakan fitur-fitur utama bibliometriks dengan mudah, meliputi: Pengumpulan data menggunakan pengumpulan dimensi, PubMed, dan Scopus API; Pemfilteran data; Analisis dan Plot; dan Analisis tiga struktur Pengetahuan (K-struktur). Biblioshiny juga terdapat dalam library pada bahasa pemrograman R sehingga tidak memerlukan untuk menginstall aplikasi lain.

Penelitian yang berkaitan dengan *literature review* pernah dilakukan sebelumnya oleh Najib *et al.* (2021) mengenai *copula model* untuk pemodelan karhutla. Dalam penelitiannya, didapatkan hasil bahwa *copula model* dapat diterapkan di berbagai bidang penelitian, termasuk analisis kebakaran hutan. Hasil *literature review* dengan metode *Systematic Literature Review* (SLR) menyajikan hasil yang terkini mengenai posisi penelitian mengenai penerapan *copula model* dalam analisis kebakaran hutan. Secara keseluruhan, penerapan metode *copula model* dalam analisis kebakaran hutan masih sangat sedikit. Namun jumlahnya semakin meningkat dan kutipan artikel untuk setiap tahun menunjukkan pentingnya penelitian ini dan telah berkontribusi terhadap pengembangan analisis karhutla.

2.3 Text Mining

Text mining, atau disebut juga dengan analisis teks, adalah sebuah teknik dalam ranah studi kecerdasan buatan yang mengubah data tidak terstruktur menjadi data terstruktur dengan menggunakan *natural language processing* (NLP) untuk meningkatkan analisis menggunakan algoritma pembelajaran mesin (Humphreys dan Jen-Hui 2018). *Text mining* sangat berguna pada masa ini karena pada platform

digital, konten-konten yang tersedia berkembang pesat dari hari ke hari dalam bentuk video, audio, gambar, dan teks. Oleh karena itu, data ini memerlukan alat dan metode yang mampu dengan mudah mengekstrak informasi berguna dari informasi tekstual (Fan *et al.* 2006). Dang dan Ahmad (2014) memandang *text mining* sebagai sebuah bidang multidisiplin, mengenai pengambilan informasi, analisis teks, ekstraksi informasi, kategorisasi, *clustering*, visualisasi, penambangan data, dan *machine learning*.

Beberapa jenis teknik *text mining* dapat diterapkan untuk menganalisis pola pada literatur teks yang besar. Beberapa contoh teknik dari *text mining* adalah klasifikasi dokumen (klasifikasi teks, standarisasi dokumen), pengambilan informasi (pencarian kata/kueri dan pengindeksan), pengelompokan dokumen (pengelompokan frasa), pemrosesan bahasa alami (koreksi ejaan, lemmatisasi, penguraian tata bahasa dan varian pengertian kata), ekstraksi informasi (ekstraksi hubungan/analisis tautan), dan analisis tautan web (Talib *et al.* 2016).

Janani dan Vijayarani (2016) menjelaskan proses *text mining* terdiri dari 6 tahap umum. Tahap *pertama*, *Document Gathering* dimana dokumen dikumpulkan dalam berbagai format. Dokumen itu dapat memproses dalam format word, html, css, dan pdf.

Kedua adalah tahap praproses dokumen. Dalam proses ini, dokumen yang tersedia akan dilakukan penghilangan redundansi, inkonsistensi, kata-kata terpisah, dan stemming. Setelah itu dokumen sudah siap untuk langkah preprocessing selanjutnya, antara lain: Tokenisasi, dimana dokumen yang diberikan dikenali sebagai string dan mengidentifikasi satu kata dalam dokumen yaitu string dokumen tertentu didistribusikan ke dalam satu unit atau token. Penghapusan *stopword*, dimana pada tahap ini telah dilakukan penghapusan kata-kata umum seperti suatu, adalah, merupakan, dan lain sebagainya. *Stemming* bertujuan untuk menghapus 2 kata yang memiliki arti yang sangat mirip.

Ketiga, *Text Transformation*, dimana dokumen teks berisi kumpulan kata dan kemunculannya. Ada dua cara untuk representasi dokumen tersebut adalah kumpulan kata-kata dan model ruang vektor. *Keempat*, *attribute selection*, dimana metode ini mengurangi ruang *database*, teknik pencarian minimal dengan menghilangkan fitur yang tidak relevan dari dokumen masukan. Ada dua metode dalam pemilihan atribut, metode *Filtering* dan *Wrapping*. *Kelima*, *pattern selection* dimana pada tahap ini proses data mining standar digabungkan dengan proses *text mining*. Penggunaan basis data terstruktur teknik *data mining* klasik yang dihasilkan dari tahap sebelumnya. Untuk mengidentifikasi pola menarik yang benar merepresentasikan pengetahuan berdasarkan beberapa prosedur yang menarik. *Keenam*, *Evaluation* dimana pada tahap ini merupakan tahapan untuk mengukur hasilnya, hasil ini dapat dibuang ataupun digunakan untuk rangkaian urutan berikutnya ataupun menjadi sebuah acuan untuk penelitian-penelitian berikutnya.

Salah satu penelitian yang berkaitan dengan *text mining* pernah dilakukan sebelumnya oleh Efendi *et al.* (2023) mengenai analisis dampak kabut asap dari karhutla di wilayah Pekanbaru serta wilayah Bengkalis, Provinsi Riau. Dalam penelitiannya, berdasarkan *term* yang muncul pada setiap hasil *cluster*, didapatkan hasil bahwa dampak yang diterima masyarakat berdasarkan hasil *clustering* yang data penelitian ini didapatkan dari media sosial Twitter yaitu masalah kesehatan serta kualitas udara.

2.4 Clustering

Clustering termasuk ke dalam teknologi *unsupervised learning* yang berarti teknologi ini berfungsi untuk mengidentifikasi pola dalam kumpulan data yang berisi titik data yang tidak diklasifikasikan atau diberi label, *clustering* bertujuan untuk mengelompokkan informasi berdasarkan kesamaan pengukuran (Zhang *et al.* 2023). *Clustering* memiliki beberapa contoh algoritma yang beragam, seperti *Regular K-Means*, *Bisecting K-Means*, dan *Unweighted Pair Group Method with Arithmetic Mean (UPGMA) Clustering*. Setiap algoritma *clustering* memiliki ciri khas masing-masing.

Algoritma *regular K-means* merupakan sebuah metode *clustering* yang umum digunakan. Algoritma ini merupakan sebuah representasi untuk algoritma *non-hierarchical clustering*. Algoritma ini digunakan dalam penelitian ini dikarenakan algoritma ini memiliki kompleksitas yang lebih mudah dan waktu proses yang lebih cepat dibandingkan algoritma *non hierarchical clustering* lainnya (Kassambara 2017).

Algoritma *Regular K Means* memiliki dua fase iterasi, yaitu fase inisialisasi yang melibatkan proses iteratif di mana setiap titik data ditetapkan ke pusat terdekat menggunakan jarak *Euclidean*, selanjutnya adalah tahap *update centroid* dimana *centroid cluster* diperbarui berdasarkan partisi yang diperoleh pada fase sebelumnya. Proses berulang berhenti ketika tidak ada titik data mengubah *cluster* atau jumlah iterasi maksimum tercapai (Oti *et al.* 2021).

Salah satu penelitian yang menggunakan algoritma *clustering* dengan K-means pernah dilakukan oleh Khairani dan Sutoyo (2020) mengenai prediksi lokasi hotspot karhutla di wilayah provinsi Kalimantan Barat, dalam penelitian ini didapatkan hasil bahwa terdapat 284 titik yang termasuk ke dalam area yang beresiko, 215 titik termasuk kedalam area yang tidak beresiko, dan 129 titik termasuk kedalam area yang sangat beresiko. Selain itu didapatkan juga hasil berupa sebuah pola untuk area yang beresiko cenderung berkumpul berdekatan, sementara area yang sangat beresiko lokasi titiknya cenderung menyebar. Gambar 1 menunjukkan proses algoritma K-Means.

- 1: Bangkitkan angka acak sebanyak k dan *centroid* k *cluster* yang sesuai;
- 2: Untuk setiap *instance*
- 3: Tetapkan *instance* ke *cluster* dengan *centroid* terdekat;
- 4: Ulangi
- 5: Perbarui *centroid*;
- 6: Tetapkan ulang *instance* ke *cluster*;
- 7: Sampai tidak ada pembaruan;

Gambar 1 Algoritma *Regular K-Means* (Banerjee *et al.* 2015)

Algoritma *Bisecting K-Means* adalah sebuah metode *clustering* yang mengkombinasikan algoritma K-Means dan *hierarchical clustering*, dengan keunggulan khusus dalam menangani data pencihan pada setiap langkah prosesnya. Algoritma ini merupakan metode *hybrid clustering* yang umum digunakan karena memiliki waktu komputasi yang lebih singkat, tingkat akurasi yang lebih tinggi, serta efisiensi yang baik ketika jumlah *cluster* bertambah (Putri *et al.* 2022). Proses utama dalam *Bisecting K-Means* adalah membelah diri, di mana titik pada *cluster*

dipecah menjadi dua centroid yang terlibat dalam proses komputasi (Zhou *et al.* 2020).

Bisecting K-Means sendiri adalah algoritma partisi yang paling umum digunakan dalam *hybrid clustering*, dimana algoritma ini mengukur jarak antar *cluster* dengan menggunakan jarak Euclidean (Banerjee *et al.* 2015). Algoritma ini bekerja dengan menginisialisasi sejumlah k buah *centroid*, lalu mengalokasikan setiap titik data ke *centroid* terdekat, dan memperbarui posisi *centroid* tersebut berdasarkan rata-rata titik data yang teralokasi pada masing-masing *centroid*. Proses ini diulang hingga konvergensi, di mana posisi *centroid* tidak lagi berubah secara signifikan.

Penelitian menggunakan algoritma *Bisecting K-Means* telah dilakukan oleh Puspitasari *et al.* (2020) dalam segmentasi pelanggan berdasarkan analisis RFM (*Recency, Frequency, Monetary*). Dalam penelitian tersebut, algoritma *Bisecting K-Means* berhasil membentuk model segmentasi pelanggan yang efektif bagi perusahaan berdasarkan data transaksi penjualan. Hasilnya menunjukkan bahwa algoritma ini mampu menghasilkan *cluster* dengan struktur medium, yang memberikan wawasan berharga bagi perusahaan dalam memahami pola perilaku pelanggan mereka. Gambar 2 menunjukkan proses algoritma *Bisecting K-Means*.

- 1: C *cluster* induk
- 2: Pilih *cluster* induk C untuk dipisah;
- 3: Ulangi
- 4: Gunakan K-Means untuk membagi C menjadi C1 dan C2;
- 5: Hitung perbedaan antar-*cluster* untuk C1 dan C2;
- 6: Hingga jumlah iterasi tetap;
- 7: Pilih sub-*cluster* dengan perbedaan antar-*cluster* tertinggi;
- 8: Hingga k *cluster* terbentuk

Gambar 2 Algoritma *bisecting K-Means* (Banerjee *et al.* 2015)

UPGMA (Unweighted Pair Group Method with Arithmetic Mean) adalah algoritma pembentukan pohon (*tree*) yang menggunakan metode berbasis jarak untuk mengelompokkan data, memanfaatkan perbedaan antara dua sekuen untuk membuat pohon tersebut (Weïß dan Göker 2011). Algoritma ini menghitung kesamaan rata-rata di seluruh titik data dalam *cluster*, menjadikannya lebih kuat dibandingkan dengan metode single linkage yang hanya mempertimbangkan titik terdekat. UPGMA menarik secara intuitif karena kemampuannya dalam menangani variasi data dengan stabilitas rata-rata aritmatika, menjadikannya algoritma yang sangat praktis dan stabil (Lowenstein *et al.* 2008).

Algoritma UPGMA adalah representasi dari algoritma *clustering* non-hierarkis, namun memiliki beberapa karakteristik yang membuatnya serupa dengan metode hierarkis. Algoritma ini sering digunakan dalam penelitian karena kompleksitasnya yang lebih sederhana dan waktu pemrosesan yang lebih cepat dibandingkan dengan algoritma *hierarchical clustering* lainnya (Kaur *et al.* 2013). Dalam UPGMA, seluruh data awalnya dianggap sebagai *cluster* tunggal. Proses *clustering* dilakukan secara berulang dengan menggabungkan pasangan *cluster* terdekat hingga terbentuk pohon *dendogram* yang lengkap. Keunggulan utama dari UPGMA adalah kemampuannya untuk memberikan struktur hierarkis yang jelas,

yang sangat berguna dalam banyak aplikasi praktis seperti filogenetik, analisis data genetik, dan pengelompokan data teks.

Contoh aplikasi untuk algoritma UPGMA dapat ditemukan dalam penelitian oleh Samuel *et al.* (2023) yang menggunakan metode ini untuk analisis dokumen publikasi terkait tingkat emisi CO dan CO₂. Penelitian ini berhasil mengelompokkan data teks menjadi beberapa *cluster* berdasarkan konstanta tingkat emisi CO dan CO₂ di Indonesia. Hasil *clustering* ini memberikan wawasan penting tentang distribusi dan pola emisi yang dapat diterapkan dalam model simulasi pergerakan kabut asap. Gambar 3 menunjukkan proses algoritma UPGMA *Clustering*.

- 1: Ulangi
- 2: Pilih perbedaan instance (D_{ij}) terkecil untuk semua pasangan $(i,j) \in S$ dengan $i \neq j$. tetapkan x sebagai parent node untuk node i dan j, hitung panjang cabang dari i ke x dan j ke x, dan catat dalam sebuah tree.
- 3: Tetapkan node baru x dengan $Dx = \frac{1}{2}(Di, Dj)$ untuk semua $k \in X$ dengan $k \neq i$ dan $k \neq j$.
- 4: Hapus i, j dari S dan tambahkan x ke dalam S.
- 5: Hingga panjang dari akar ke masing-masing daun pada kedua cabang mencapai n-1

Gambar 3 Algoritma UPGMA (Yoshida *et al.* 2022)

2.5 Bibliometriks

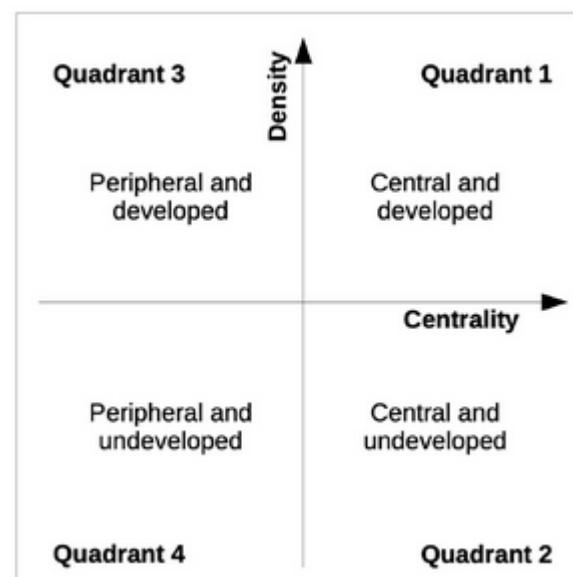
Analisis bibliometriks adalah metode yang populer dan sangat baik untuk mengeksplorasi dan menganalisis sejumlah besar data untuk kebutuhan ilmu pengetahuan. Hal ini memungkinkan untuk mendefinisikan berbagai macam pola yang dicari pada suatu bidang tertentu dengan cara melihat bidang atau tema-tema yang sedang berkembang wilayah di bidang tersebut (Donthu *et al.* 2021)

Teknik untuk analisis bibliometriks terbagi ke dalam dua kategori, yaitu kategori analisis kinerja dan analisis pemetaan sains. Analisis kinerja akan memperhitungkan kontribusi komponen penelitian dari pengguna, sedangkan analisis pemetaan sains berfokus pada hubungan antar elemen konstituen penelitian (Donthu *et al.* 2021). Bibliometriks juga dapat digunakan untuk mengidentifikasi mitra-mitra utama, melalui survei mengenai siapa kontributor yang paling banyak dalam kumpulan dokumen teks artikel publikasi. Sebagai administrator penelitian, pengetahuan dasar tentang terminologi dan alat bibliometriks adalah penting, dan hubungan yang baik dengan perpustakaan atau kelompok analitik organisasi juga sama pentingnya (Andersen 2018).

Salah satu *tools* atau fungsi dalam analisis bibliometriks adalah peta tematik (*thematic map*) dan pemetaan kuadran. Peta tematik digunakan untuk memvisualisasikan dan mengklasifikasikan berbagai tema atau topik penelitian berdasarkan relevansi dan kekuatannya dalam dunia literatur yang sedang berkembang serta memvisualisasikan keterhubungannya antara satu tema dengan tema lainnya. Kuadran pada peta tematik membagi tema berdasarkan dua dimensi utama, yaitu pentingnya tema dalam bidang penelitian (*centrality*) dan tingkat pengembangan atau kedewasaan tema tersebut (*density*) sehingga menghasilkan 4



kuadran yang merepresentasikan dimensi-dimensi tadi. Kuadran 1 (*Central and developed*) yang merupakan quadran *high centrality* dan *high density* berisi tema yang sangat penting dan berkembang baik, menjadi pusat perhatian penelitian. Kuadran 2 (*Central and undeveloped*) yang berisi *high centrality* dan *low density* merupakan dari tema mendasar yang penting untuk berbagai penelitian lain tetapi mungkin tidak memiliki kedalaman penelitian yang besar. Kuadran 3 (*Peripheral and developed*) dengan *high density* dan *low centrality* mencakup tema yang sangat berkembang tetapi kurang terhubung dengan tema lain, seringkali sangat spesifik. Kuadran 4 (*Peripheral and undeveloped*) dengan *low centrality* dan *low density* menampilkan tema yang baru muncul atau menurun, yang masih dalam tahap awal atau kehilangan popularitas (Aria *et al.* 2020). Visualisasi untuk tiap-tiap kudaran dapat dilihat pada Gambar 4



Gambar 4 Kuadran thematic map dalam analisis bibliometriks (Yu dan Munoz 2020)

2.6 Linkage

Tipe dasar dalam metode hierarki atau *linkage* dapat berupa aglomeratif atau devisif. Pada metode aglomeratif, dimulai dengan menempatkan obyek dalam *cluster-cluster* yang berbeda kemudian mengelompokkan obyek secara bertahap ke dalam *cluster-cluster* yang lebih besar, sedangkan pada *cluster* devisif dimulai dengan menempatkan semua obyek sebagai satu *cluster*. Kemudian secara bertahap obyek-obyek dipisahkan ke dalam *cluster-cluster* yang berbeda, dua *cluster*, tiga *cluster*, dan seterusnya (Simamora 1996). Terdapat 5 jenis metode *linkage clustering*, yaitu *Single linkage*, *Complete Linkage*, *Average linkage*, *Ward's method*, dan *Centroid*

2.4.1 Single Linkage

Metode *single linkage* adalah proses pengklasteran yang didasarkan pada jarak terdekat antar objek. Jika dua objek terpisah oleh jarak yang pendek, maka kedua objek tersebut akan digabung menjadi satu *cluster*. Metode ini dimulai dengan menentukan objek-objek yang memiliki jarak terkecil dalam matriks

proximity D = {d_{ik}}. Objek-objek yang memiliki jarak terkecil digabungkan dalam satu *cluster*, misal A dan B adalah dua objek yang akan diklasterkan sehingga diperoleh *cluster* (AB) (Wahyuni *et al.* 2013). Untuk mencari jarak antara *cluster* (AB) dan *cluster* W atau *cluster* lainnya diperoleh pada Persamaan 1

$$d(AB)W = \min \{dAW, dBW\} \quad (1)$$

dengan:

- | | |
|--------|---|
| d(AB)W | = jarak minimum antara kelompok (AB) dan kelompok W |
| dAW | = jarak terdekat dari <i>cluster</i> A dan W |
| dBW | = jarak terdekat dari <i>cluster</i> B dan W |

2.4.2 Complete Linkage

Complete Linkage disebut dengan *maksimum link*, dimana jarak antar dua buah *cluster* dihitung dari jarak terjauh antara anggota *cluster* yang satu dengan *cluster* yang kedua. Kelebihannya terletak pada sedikit pengaruh terhadap *noise* dan *outliers*, sedangkan kekurangannya adalah cenderung memecah *cluster* dengan ukuran yang besar dan lebih mengarah pada sekumpulan *cluster* berbentuk bulat. (Damayati *et al.* 2023). Adapun perhitungan pada rumus metode *Complete Linkage clustering* disajikan dalam Persamaan 2

$$d(AB)W = \max \{dAW, dBW\} \quad (2)$$

dengan:

- dAW adalah jarak terdekat dari klaster A dan W
 dBW adalah jarak terdekat dari klaster B dan W

2.4.3 Average Linkage

Metode *average linkage* adalah metode *clustering* dengan prinsip jarak rata-rata antar setiap pasangan objek yang mungkin pada satu *cluster* dengan seluruh objek pada *cluster* yang lain. *Average linkage* menghitung jarak antara dua *cluster* yang disebut sebagai jarak rata-rata dimana jarak tersebut dihitung pada masing-masing *cluster*. Rumus untuk metode *average linkage* dapat dilihat pada Persamaan 3 (Johnson dan Wichern 2007)

$$d_{(ab)w} = \frac{\sum_i \sum_k d_{ik}}{N_{AB} N_w} \quad (3)$$

dengan:

- | | |
|-----------------|--|
| d _{ik} | = jarak antar objek i dalam <i>cluster</i> (AB) dan objek k dalam <i>cluster</i> W |
| N _{AB} | = jumlah item pada <i>cluster</i> AB |
| N _w | = jumlah item pada <i>cluster</i> W |

2.4.4 Ward's method

Metode *Ward* adalah metode *clustering* yang bersifat *agglomerative* (penggabungan) untuk memperoleh kelompok yang memiliki varian *internal*

sekecil mungkin. Metode ini menggunakan perhitungan yang lengkap dan memaksimumkan homogenitas di dalam satu kelompok. Untuk *clustering* metode *Ward*, jumlah kelompok ditentukan terlebih dahulu berdasarkan dendogram. Ukuran yang digunakan adalah *Sum of Square Error* (SSE) *variable*. *Agglomerative* merupakan prosedur pengelompokan hierarki dimana setiap objek berasal dari *cluster* yang terpisah. *Cluster-cluster* dibentuk dengan mengelompokkan objek ke dalam *cluster* yang semakin banyak objek yang menjadi anggotanya. Proses ini dilanjutkan sampai semua objek menjadi anggota dari *cluster* tunggal. Metode *Ward* merupakan bagian dari metode pengelompokan yang mengelompokkan N buah objek ke dalam $N-1$, $N-2$, dan seterusnya sampai menjadi satu *cluster*, dengan banyaknya *cluster* tidak diketahui. SSE hanya dapat dihitung jika *cluster* memiliki elemen lebih dari satu objek. SSE *cluster* yang hanya memiliki satu objek adalah nol. Metode *Ward* dengan perhitungan SSE ini dihitung berdasarkan Persamaan 4 (Oktavia *et al.* 2013):

$$SSE = \sum_{i=1}^n (X_i - \bar{X})' - (X_i - \bar{X}) \quad (4)$$

dengan:

X_i = vector kolom yang entrinya nilai objek ke-I dengan $i= 1,2,3,\dots,N$

\bar{X} = vector kolom yang entrinya rata-rata nilai objek dalam *cluster*

N = banyaknya objek dalam *cluster* yang terbentuk

2.4.5 Centroid

Metode ini menitikberatkan pada nilai rata-rata semua obyek dalam *cluster*. Jarak yang digunakan dalam proses identifikasi *cluster* adalah kedekatan dengan titik *centroid* *cluster* yang terbentuk. *Cluster centroid* yang terbentuk nantinya adalah nilai tengah observasi pada variabel dalam suatu *set* variabel *cluster*. Dengan metode ini, setiap terjadi klaster baru segera terjadi perhitungan ulang *centroid* sampai terbentuk *cluster* yang cenderung konvergen. Keuntungan dari metode ini adalah tidak berpengaruh secara signifikan terhadap pencarian dibandingkan dengan metode lain (Nielsen 2016), Persamaan 5 menunjukkan rumus untuk metode *centroid*.

$$[\{X_n\}, \{Y_n\}] = \left\| \left(\frac{1}{N} \sum_{n=1}^N X_n \right) - \left(\frac{1}{M} \sum_{m=1}^M Y_m \right) \right\| \quad (5)$$

dengan :

N : jumlah semua observasi

X_n : Objek X ke-n

Y_n : Objek Y ke-n

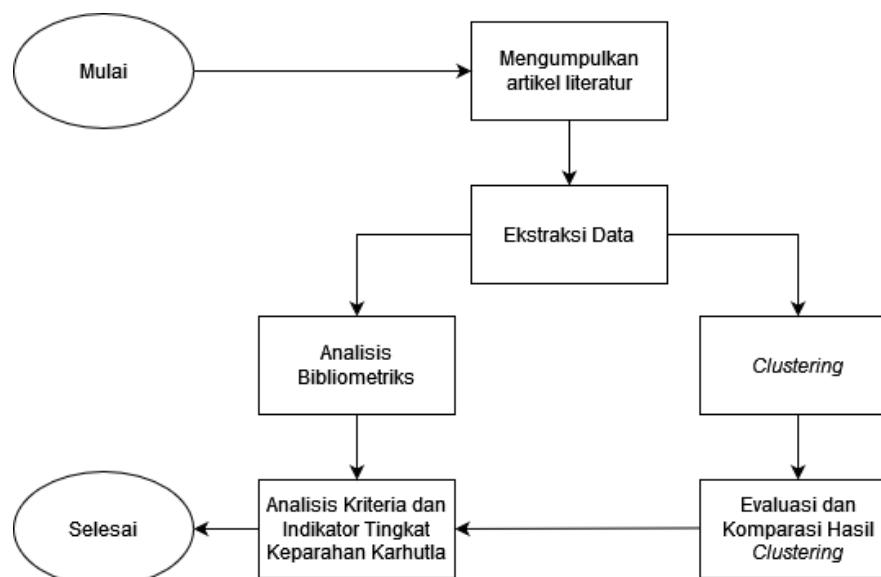
III. METODE

3.1 Data Penelitian

Data penelitian yang digunakan dalam penelitian ini berupa literatur jurnal, konferensi, atau publikasi terakreditasi terindeks scopus dengan semua quartil dan jurnal nasional terindeks Sinta 1, 2 dan 3. Data literatur paper dikumpulkan berdasarkan kata kunci pencarian. Data artikel publikasi didapatkan dari *database* Google Scholar dan disimpan dalam aplikasi Mendeley Reference Manager dan secara lokal juga. Literatur yang akan dianalisis dapat bersumber dari dalam negeri ataupun luar negeri. Artikel publikasi yang digunakan dalam penelitian ini bersifat *open access*.

3.2 Tahapan Penelitian

Tahapan penelitian ini terdiri atas beberapa tahapan yang disajikan pada Gambar 5. Penelitian ini terdiri atas 6 tahapan seperti dijelaskan berikut ini:



Gambar 5 Diagram alur penelitian

3.2.1 Mengumpulkan artikel literatur sesuai dengan kata kunci pencarian

Pada tahap ini, artikel literatur akan dikumpulkan dengan menggunakan *software* publish or perish dengan kata-kunci pencarian seperti *forest fire*, *fire severity*, *criteria*, *indicator*, dan *indonesia*. Artikel literatur yang dikumpulkan berupa artikel publikasi jurnal maupun konferensi yang dipublikasikan pada tahun 2013 hingga 2023. Artikel literatur dapat berupa jurnal ataupun paper konferensi. Literatur-literatur yang akan dianalisis dapat bersumber dari dalam dan luar negeri. Literatur yang dikumpulkan juga akan diekstrak dalam bentuk data *text* dengan pendekatan *text mining* untuk dikelompokkan pada tahap-tahap selanjutnya.



3.2.2 Ekstraksi Data Hasil dari *Text Mining*

Artikel yang sudah dikumpulkan akan dilakukan proses ekstraksi data *text* dimana data yang diekstrak merupakan data *full text* yang diambil dari artikel publikasi. Hasil ekstraksi artikel publikasi yang telah dikumpulkan akan ditampilkan dengan menggunakan *cloud word*. Hasil ekstraksi dari data *full text* akan dilakukan *text preprocessing* untuk memilih kata-kata yang akan dijadikan input untuk tahap berikutnya. Praproses akan melalui beberapa tahapan, yang pertama adalah *case folding* dimana tujuannya adalah untuk menghilangkan *uppercase* dari *data text* yang ada, kemudian tahap *tokenizing* dimana tujuannya adalah untuk menghilangkan tanda baca, simbol, angka, dan *whitespace*, kemudian tahapan dilanjutkan dengan *filtering* dimana tahap ini akan menghilangkan kata-kata penghubung, tahap terakhir adalah *stemming* dimana tahap ini akan menghilangkan kata-kata imbuhan pada kata yang akan dianalisis.

Hasil ekstraksi setelah dilakukan praproses akan ditabulasikan berdasarkan judul. Karakteristik yang akan digunakan berupa *term* yang berkaitan indikator keparahan kebakaran hutan. *Term* karakteristik hasil ekstraksi akan digunakan sebagai data input pada tahap *clustering*.

3.2.3 Analisis Bibliometriks

Artikel yang telah berhasil dikumpulkan dengan *software publish or perish* akan menjalani analisis terkait dengan pola keterhubungan dan interpretasi antara artikel. Proses analisis ini akan dilakukan melalui pendekatan bibliometriks dengan memanfaatkan perangkat lunak *Biblioshiny*, dimana software ini merupakan sebuah alat visualisasi yang secara khusus dirancang untuk menggambarkan dan memahami jaringan kolaborasi serta struktur literatur ilmiah. Analisis bibliometriks bertujuan untuk melihat bagaimana pola keterhubungan antara satu sumber penelitian dengan penelitian lainnya. Selain itu analisis bibliometriks ini juga dapat melihat relevansi tema penelitian terkait yang diambil dari dataset teks publikasi.

3.2.4 *Clustering*

Pada tahap ini, hasil ekstraksi data teks yang didapatkan pada tahap sebelumnya akan digunakan sebagai input untuk tahap *clustering*. Pemodelan *cluster* akan menggunakan 3 metode *clustering*, yaitu *Regular K-Means*, *Bisecting K-Means*, dan *UPGMA Clustering*. *Regular K-Means* merepresentasikan *non-hierarchical clustering*, *UPGMA* merepresentasikan *hierarchical clustering*, dan *Bisecting K-Means* merepresentasikan pendekatan *hybrid*. Metode tersebut dipilih berdasarkan metode yang umum dan mudah digunakan.

3.2.5 Evaluasi dan Komparasi Metode *clustering*

Hasil *clustering* dengan ketiga metode *clustering* yang telah dilakukan pada tahapan sebelumnya kemudian akan dievaluasi menggunakan perhitungan *silhouette score*. Rentang dari nilai *sillhouette score* adalah antara 1 sampai -1, dimana banyaknya *cluster* dengan nilai *silhouette score* yang mendekati 1 pada setiap metode *clustering* yang digunakan menggambarkan jumlah *cluster* yang optimum atau paling baik. Perhitungan *silhouette score* dapat dilihat pada Persamaan 6 (Shutawyi dan Kachouie 2021).

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (6)$$

dengan:

- s = *silhouette score*
- a = rataan jarak antar data pada *cluster* yang berbeda
- b = rataan jarak antar titik/data dalam *cluster* yang sama

Nilai *silhouette score* yang bernilai dibawah 0 atau negatif dapat dikatakan merupakan nilai *silhouette score* yang tidak baik karena $a(i)$ lebih besar dari $b(i)$, yang berarti ketidaksamaan dalam kelompok *cluster* lebih besar daripada ketidaksamaan antar kelompok *cluster*. Nilai positif diperoleh ketika $a(i) < b(i)$, dan lebar *Silhouette score* mencapai maksimum $s(i) = 1$ ketika $a(i) = 0$. Semakin besar nilai $s(i)$ yang positif, semakin besar kemungkinan elemen tersebut dikelompokkan dalam kelompok *cluster* yang benar. Elemen atau anggota *cluster* dengan nilai $s(i)$ negatif lebih mungkin termasuk dalam kelompok *cluster* yang salah (Shutawyi dan Kachouie 2021).

3.2.6 Analisis Kriteria dan Indikator Tingkat Keparahan Karhutla

Hasil analisis dengan *clustering* dan analisis bibliometriks dianalisis pada tahap ini, hasil *clustering* akan dianalisis secara manual untuk mencari kriteria dan indikator keparahan karhutla. Analisis dengan bibliometriks juga akan dilakukan untuk melihat persamaan hasil antara hasil dari analisis *clustering* dan hasil dari analisis bibliometriks

3.3 Lingkungan Pengembangan

Spesifikasi perangkat keras dan perangkat lunak yang digunakan untuk penelitian ini adalah sebagai berikut:

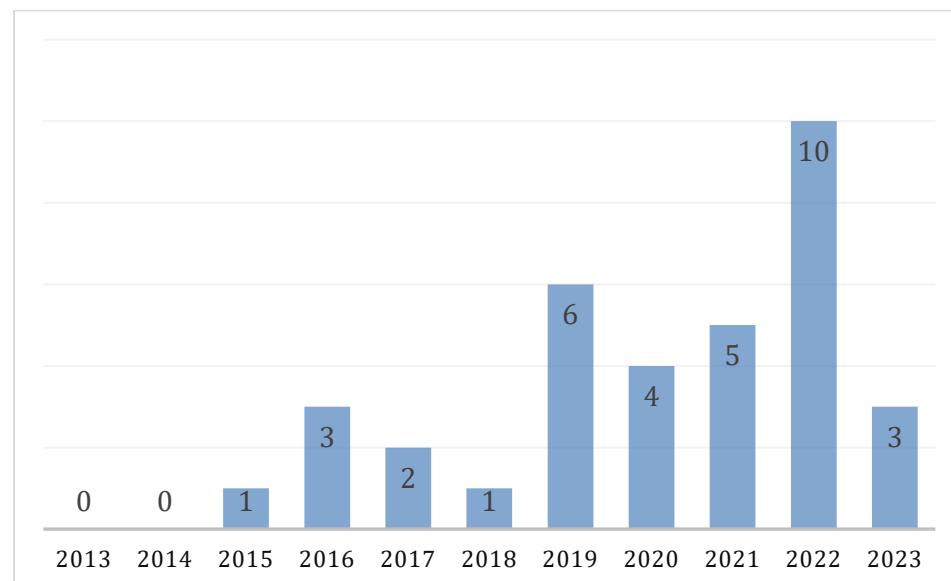
1. Perangkat keras yang digunakan berupa laptop personal dengan spesifikasi:
 - Processor Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz 1.80 GHz
 - Storage 225 GB
 - RAM 16 GB
2. Spesifikasi perangkat lunak yang digunakan:
 - Sistem Operasi Windows 10
 - Rstudio 1.4.1717 untuk melakukan proses *clustering*
 - Publish or perish 8 untuk mengumpulkan artikel literatur
 - Biblioshiny untuk analisis bibliometriks

IV. HASIL DAN PEMBAHASAN

4.1 Mengumpulkan Artikel Literatur

Artikel literatur dikumpulkan menggunakan aplikasi manajemen referensi berupa aplikasi Publish or Perish dan website Scopus. *Keyword* yang digunakan dalam penelitian ini adalah *forest fire*, *fire severity*, *criteria*, *indicator*, dan *Indonesia*. Dimana kata kunci *forest fire* dan *fire severity* dipilih dikarenakan kesesuaian dengan tema penelitian, kata kunci *criteria* dan *indicator* dipilih dikarenakan kedua kata ini merupakan bagian dalam penelitian, dan Indonesia yang merupakan salah satu ruang lingkup dalam penelitian ini. Referensi artikel literatur dikumpulkan dengan rentang waktu 2013 hingga 2023 dengan frekuensi seperti dalam Gambar 6.

Hasil penelusuran disimpan dalam berkas dengan format csv dan ris. Berkas-berkas artikel literatur diunduh dengan aplikasi Publish or Perish menggunakan rujukan *file* pada hasil dari pencarian sebelumnya. Pencarian melalui Publish or Perish dapat dilihat dalam Gambar 7. Artikel yang berhasil dikumpulkan dengan aplikasi publish or perish mencapai hingga 300 artikel publikasi. Artikel publikasi yang sudah terkumpul lalu dilakukan *filtering* dengan parameter berupa akreditasi artikel dengan level akreditasi sinta 1, 2, 3 atau scopus, lalu dilakukan kembali *filtering* untuk artikel publikasi dengan status *open access*. Setelah dilakukan tahap *filtering* untuk pengumpulan data, artikel kembali di filtering untuk mengecek kesesuaian artikel dengan tema penelitian. Setelah dilakukan filtering, jumlah artikel publikasi yang digunakan dalam penelitian ini berjumlah sebanyak 35 artikel. Artikel-artikel ini dapat dilihat dalam Lampiran 1.



Gambar 6 Frekuensi artikel per tahun yang digunakan

The screenshot displays two search interfaces side-by-side. On the left is a Google Scholar search results page with various filters applied. On the right is a 'Publish or Perish' tool interface showing a detailed list of publications related to the search terms.

Cites	Per year	Rank	Authors	Title	Year	Publication	Publisher	Type
11	5.50	1	T Nikonova, A Spe...	ProbFire: a probabilistic fire early ...	2022	Hazards and Earth ...	nhes.copernicus.org	HTML
40	5.71	2	E Sumarga	Spatial indicators for human activi...	2017	Tropical Conservation Scie...	journals.sagepub.com	HTML
10	2.50	3	Ri Agus, C Ahmad, ..	Understanding forest fire manageme...	2020	ASEAN Journal on ...	aistdubd.edu.bn	HTML
13	4.33	4	AA Fitriany, PJ Flata...	Assessment on the use of meteorolo...	2021	Sustainability	mdpi.com	HTML
h 84	14.00	5	A Ibar, INS Jaya, RH...	Spatio-temporal analysis of land ...	2018	Jurnal Wilayah Dan Lingku...	academia.edu	PDF
4	1.33	6	MBR Prayoga, RH...	Improving forest fire mitigation in...	2020	Remote Sensing of ...	elsevier	HTML
h 110	27.50	7	L Liu, LJ Mickey, M...	Diagnosing spatial biases and inc...	2019	Environmental Research Le...	ippscience.iop.org	HTML
36	7.20	8	JS Sze, Jefferson, JS...	Evaluating the social and environ...	2015	Procedia Environmental Sc...	elsevier	HTML
30	3.33	9	A Nurdiana, I Ridi...	Indicator determination of forest ...	2021	... series: Earth and ...	ippscience.iop.org	HTML
5	1.67	10	DE Nurianto, RP Pr...	Developing models to establish s...	2018	Forests	mdpi.com	HTML
19	3.17	11	A Krassowski, N Kha...	Modeling burned areas in Indonesia...	2012	World Development	elsevier	HTML
h 110	27.50	12	RB Edwards, RL Na...	Causes of Indonesia's forest fires	2017	Biogeosciences	bg.copernicus.org	HTML
74	10.57	13	T Sanin, GR Von Do...	Precipitation-fire linkages in Indo...	2019	Forests	mdpi.com	HTML
38	7.60	14	HA Adrianto, DV S...	Relationship between fire and for...	2019	Journal of Science and Tec...	emerald.com	HTML
15	3.00	15	AAR Fernandes, RB...	The effect of community and com...	2018	Integrating Disaster Scienc...	elsevier	HTML
38	6.33	16	L Syaulina	Forest and land fires in Indonesia...	2015	... Hazards and Earth ...	nhes.copernicus.org	HTML
h 76	8.44	17	AC Spessa, RD Fiet...	Seasonal forecasting of fire over K...	2016	Indonesian Journal of ...	academia.edu	PDF
13	1.63	18	A Ibar, INS Jaya, B...	Spatio-temporal typology of land ...	2016	... and characterization of low-tem...	journals.plos.org	HTML
h 73	9.13	19	EC Atwood, S Engl...	Forestry, forest fires, and climate c...	2017	Bulletin of Indonesian Eco...	Taylor & Francis	HTML
h 139	19.86	20	AS Alisahabana, JM...	Foresty, forest fires, and climate c...	2019	International Forestry ...	ingentaconnect.com	HTML
57	11.40	21	R Putromo, B Okar...	Forest and land fires, toxic haze ...	2011			

Gambar 7 Penelusuran dengan Publish or Perish

4.2 Ekstraksi Data

Kumpulan artikel yang sudah dikumpulkan dalam proses sebelumnya akan diekstraksi dengan metode *text mining* untuk mendapatkan *term-term* yang akan digunakan pada tahapan proses *clustering*. Tahapan proses *text mining* dimulai dengan membaca semua *pdf file* dengan fungsi Corpus. Hasil corpus dipilah dengan mengecilkan huruf pada dokumen, kemudian dilanjutkan dengan menghapus angka, tanda baca, dan sebagainya dengan fungsi TermDocumentMatrix. Contoh dari *term* dan frekuensi yang muncul dari seluruh artikel yang telah diproses dapat dilihat dalam Tabel 2 dan kemudian dipetakan dalam *wordcloud* yang dapat dilihat dalam Gambar 8. Hasil *term* dan frekuensi yang muncul dapat dilihat dalam Gambar 9.

Tabel 2 *Term* dan frekuensi 10 teratas

No	Term	Frekuensi
1	<i>forest</i>	9205
2	<i>area</i>	8339
3	<i>burn</i>	5946
4	<i>use</i>	5482
5	<i>land</i>	5391
6	<i>data</i>	5296
7	<i>fire</i>	5280
8	<i>hotspot</i>	4116
9	<i>sever</i>	3607
10	<i>peatland</i>	3542



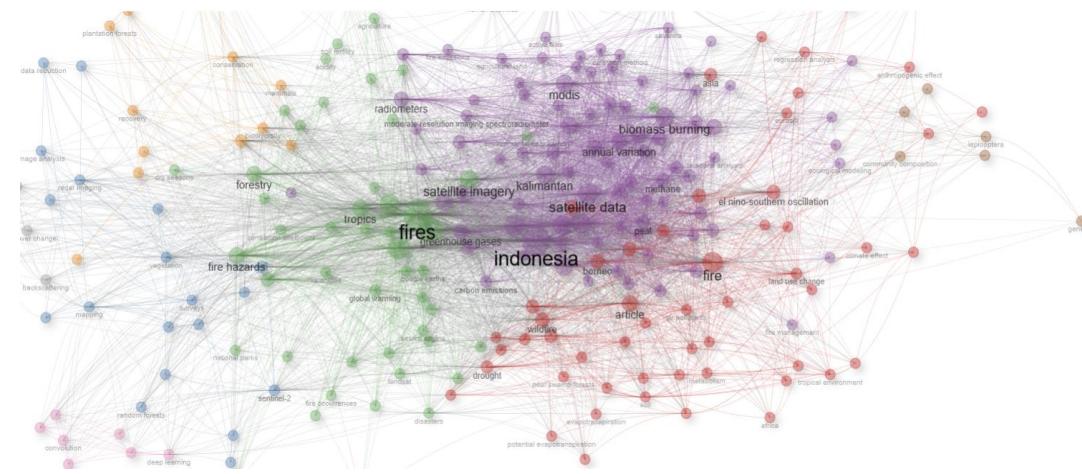
Gambar 8 Wordcloud kumpulan artikel literatur

▲	abil	▼	abl	▼	absolut	▼	abstract	▼	academ	▼	accept	▼	access	▼	accord	▼	account	▼	accumul	▼	accur	▼	accuraci	▼
1	1		2		1		23		7		24		27		10		3		4		6		15	
2	2		1		0		9		0		4		11		15		9		4		7		14	
3	3		3		0		4		1		0		18		7		2		2		5		20	
4	2		1		7		1		0		0		14		9		3		2		3		19	
5	7		3		1		0		0		0		5		10		5		8		0		22	
6	4		3		1		0		0		2		7		7		8		9		2		42	
7	2		1		0		0		0		2		5		6		4		4		7		19	
8	1		1		2		0		0		0		16		6		5		1		1		7	
9	8		0		3		1		3		1		4		9		5		6		4		17	
10	0		0		1		3		2		4		7		9		1		10		3		24	
11	0		3		0		1		1		4		10		15		3		9		5		21	
12	5		5		1		2		0		1		16		9		3		3		12		18	
13	3		1		2		4		2		4		20		10		5		1		9		25	
14	1		4		0		3		1		3		4		4		4		1		6		28	

Gambar 9 Hasil *text extraction*

4.3 Analisis Bibliometriks

Dalam analisis bibliometriks menggunakan fungsi *thematic map*, kata-kata yang berkaitan akan dihubungkan antara satu sama lainnya. *Thematic map* menunjukkan adanya beberapa keterkaitan antara 8 *cluster* atau kelompok tema penelitian yang secara otomatis sudah dihitung oleh Biblioshiny dan divisualisasikan bedasarkan perbedaan warna pada gambar 10. Kelompok tema yang dihasilkan pada *thematic map* adalah Indonesia, *peatland, satellite data; Fire, carbon, Borneo; fire, deforestation, Burned area; Synthetic aperture radar, mapping, radar imaging; ecosystem, conservation, biodiversity; Convolution, deep learning, CNN; lepidoptera, butterfly, community composition; dan backscattering, land-cover change, operational land imager*. Selain itu, *thematic map* juga menunjukkan frekuensi tema artikel yang banyak muncul untuk tiap *cluster* tema, hal ini dapat dilihat pada Gambar 10 juga yang memperlihatkan adanya perbedaan ukuran pada ukuran lingkaran pada tema yang divisualisasikan, semakin besar lingkaran tersebut, maka frekuensi artikel dengan kata kunci berkaitan semakin besar. frekuensi tema penelitian yang muncul dari seluruh artikel yang telah diproses di Biblioshiny dapat dilihat dalam Tabel 3.



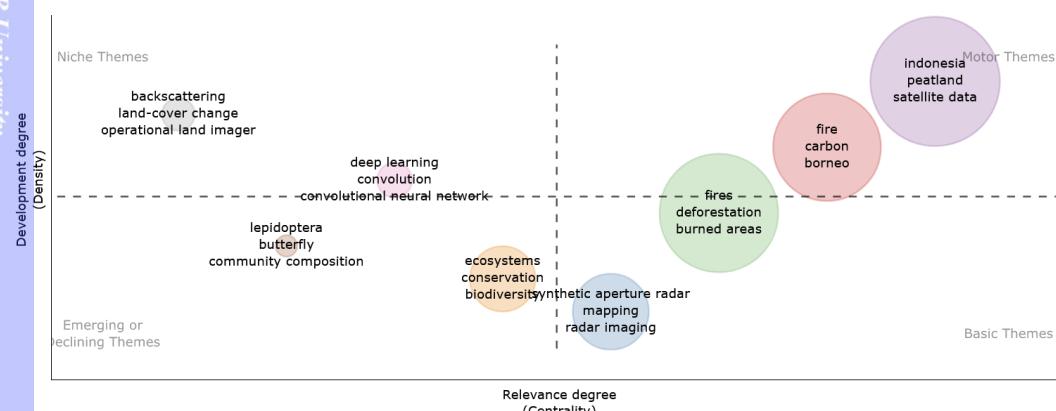
Gambar 10 Hasil *thematic map network* dengan menggunakan Biblioshiny

Tabel 3 Frekuensi tema penelitian teratas

No	Tema penelitian	Frekuensi
1	<i>fires</i>	47
2	<i>Indonesia</i>	45
3	<i>deforestation</i>	26
4	<i>burned area</i>	22
5	<i>Remote sensing</i>	20
6	<i>Fire hazard</i>	16
7	<i>forestry</i>	15
8	<i>peatland</i>	15
9	<i>Satellite image</i>	14
10	<i>Satellite data</i>	13

Thematic map yang merupakan bagian dari analisis bibliometriks menghasilkan empat buah kuadran untuk menunjukkan seberapa pentingnya suatu tema dalam bidang penelitian (*centrality*) dan tingkat pengembangan atau tingkat kedewasaan tema tersebut (*density*). Dalam kuadran yang dihasilkan, dapat diketahui bahwa pada kuadran *motor theme* (kuadran 1) yang merupakan kuadran dengan jumlah penelitian dan keterhubungan yang besar, terdapat tema *1 Indonesia, peatland, satellite data*, dan juga tema *fire, carbon, Borneo*, dengan tema *fire, deforestation, Burned area* yang beririsan dengan kuadran 2. Pada kuadran *basic* (kuadran 2) yang merupakan kuadran dengan keterhubungan yang besar tetapi dengan jumlah penelitian yang sedikit, terdapat tema *Synthetic aperture radar, mapping, radar imaging* dan juga terdapat tema *fire, deforestation, Burned area* yang beririsan dengan tema 1. Pada kuadran *emerging or declining* (kuadran 3) yang merupakan kuadran dengan keterhubungan dan jumlah penelitiannya sedikit, terdapat tema *lepidoptera, butterfly, community composition, and tema ecosystem, conservation, biodiversity*. Pada kuadran *niche* (kuadran 4) terdapat tema *Convolution, deep learning, CNN, and juga tema backscattering, land-cover change, operational land imager*.

Motor theme (kuadran 1) dan *basic theme* (kuadran 2) memiliki *relevance degree* tinggi yang menandakan bahwa tema-tema yang terdapat pada kedua kuadran tersebut banyak digunakan oleh penelitian dalam artikel literatur yang diambil dalam jangka waktu 10 tahun terakhir. Pada Gambar 11 dapat dilihat terdapat 8 buah *cluster* yang sudah dikelompokan secara otomatis oleh Biblioshiny dengan 4 buah kuadran yang telah dibagi-bagi, dapat dilihat juga terdapat lingkaran dengan berbagai ukuran pada tiap-tiap *cluster*. Hal ini dapat diartikan bahwa dalam setiap *cluster*, semakin banyak tema yang merupakan anggota dari suatu *cluster*, maka lingkaran pada *thematic map* akan semakin besar. Visualisasi tema yang merupakan anggota tiap *cluster* pada Gambar 11 merupakan 3 tema dengan frekuensi terbesar untuk setiap *cluster*.



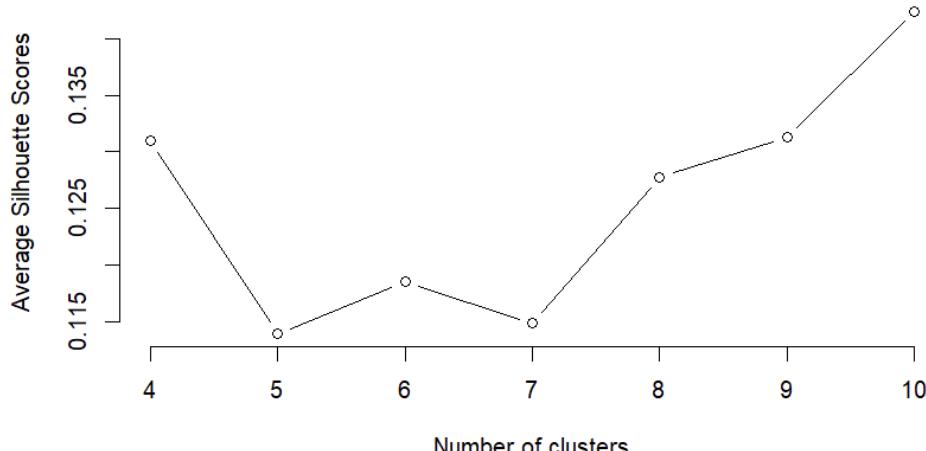
Gambar 11 pembagian kuadran hasil analisis bibliometriks beserta ukuran untuk setiap *cluster*

4.4 Clustering

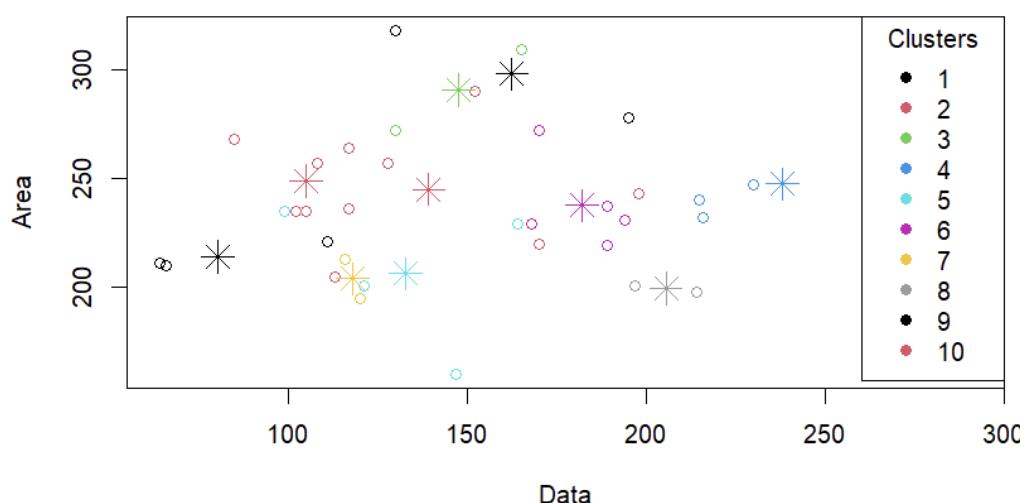
Tahap pelaksanaan metode *Clustering* dilakukan dengan menginput *term-term* yang merupakan hasil dari 35 artikel literatur yang telah diproses pada tahapan *text mining* sebelumnya. *Clustering* artikel literatur dilakukan dengan 3 metode algoritma, yaitu *bisecting K-Means*, *regular K-Means*, dan *UPGMA clustering*. Algoritma *bisecting K-Means* dibangun dengan *library cluster* dengan bahasa pemrograman R. Algoritma *regular K-Means* dibangun dengan *library cluster* dengan bahasa pemrograman R. Algoritma *UPGMA clustering* dibangun dengan *library cluster* dan *hclus* dengan bahasa pemrograman R.

4.4.1 Clustering dengan algoritma *regular K-Means*

Clustering dengan algoritma *regular K-Means* dilakukan dengan bahasa pemrograman R. Data yang diperoleh pada tahap sebelumnya dimuat ke dalam sebuah *dataframe*. Data *input* diproses dengan fungsi yang dibuat untuk mendapatkan nilai dari *silhouette score* dengan parameter input fungsi berupa k buah *cluster* yang akan diproses. Jumlah *cluster* yang diproses adalah sebanyak 4 hingga 10 buah *cluster*. Jumlah *cluster* yang terbaik ditunjukkan pada Grafik *silhouette score Regular K-means* dalam Gambar 12 yang menunjukkan *cluster* terbaik adalah sebanyak 10 *cluster*

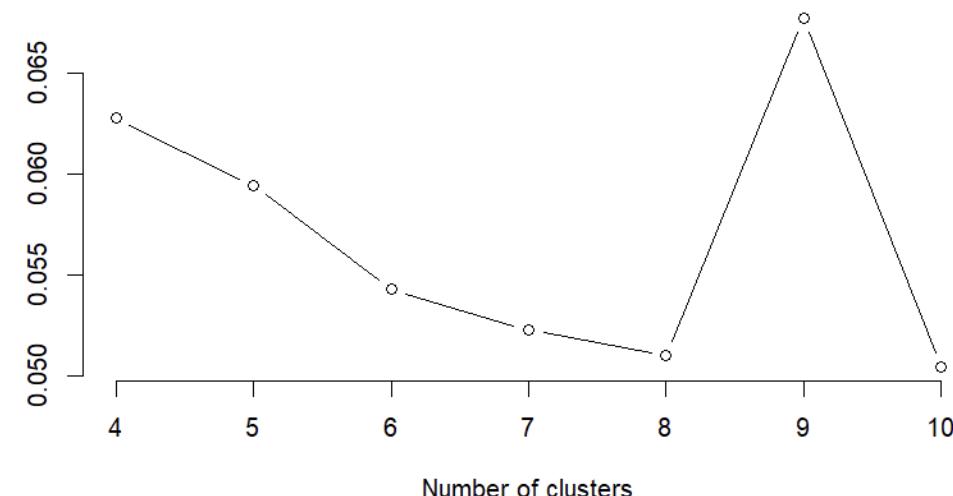
Gambar 12 *Silhouette score* dari algoritma *regular K-Means*

Silhouette score yang dihasilkan algoritma *regular K-Means* adalah 0,130; 0,113; 0,118; 0,126; 0,127; 0,131; dan 0,138. Berdasarkan hasil tersebut, dapat disimpulkan bahwa jumlah *cluster* yang terbaik untuk dataset yang digunakan dalam penelitian ini adalah 10 *cluster*, dengan nilai *silhouette score* tertinggi sebesar 0,138. Hal ini menunjukkan bahwa pembagian data ke dalam 10 *cluster* memberikan hasil yang paling optimal, dengan tingkat keseragaman internal *cluster* yang tinggi dan pemisahan antar *cluster* yang jelas. Sebaran *cluster* yang dihasilkan oleh algoritma *Bisecting K-Means* dipetakan dalam fungsi *plot* dengan dimensi yang ditampilkan berupa kolom data dan area. Hasil pemetaan untuk sebaran *cluster* ini memberikan gambaran visual mengenai pola distribusi data dalam ruang fitur yang diwakili oleh nilai dari *term forest* dan *area*. Visualisasi ini membantu memahami bagaimana data terbagi ke dalam *cluster* yang berbeda serta karakteristik dari *cluster*. Gambar 13 menunjukkan 10 *cluster* yang telah terbagi dengan algoritma *regular K-Means* dengan contoh visualisasi *cluster* adalah term *Area* dan term *Data* dimana nilainya adalah jumlah frekuensi *term* yang terdapat pada *dataset*. Anggota-anggota dari tiap *cluster* dapat dilihat pada Lampiran 2.

Gambar 13 Sebaran *cluster* dari algoritma *regular K-Means*

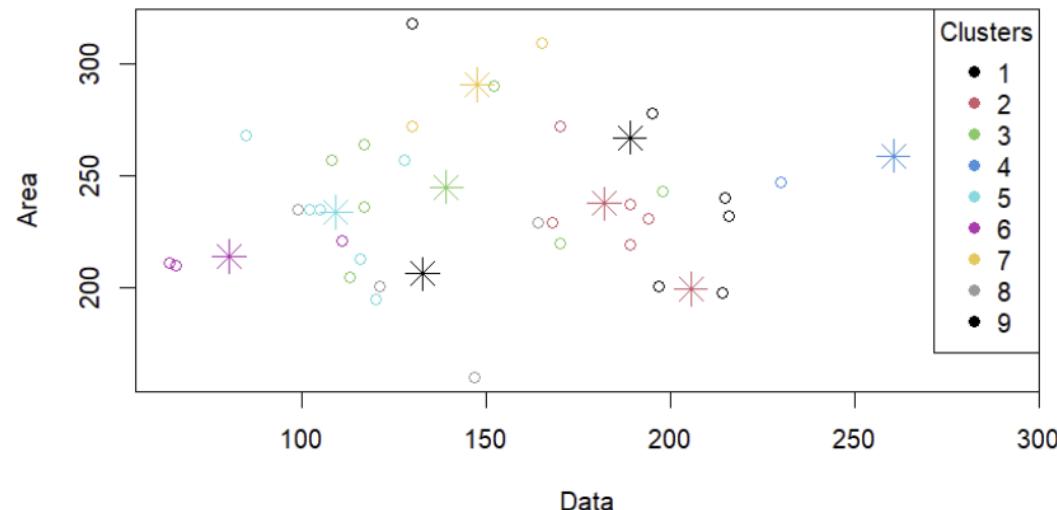
4.4.2 Clustering dengan algoritma Bisecting K-Means

Clustering dengan algoritma Bisecting K-Means dilakukan menggunakan bahasa pemrograman R. Data yang diperoleh pada tahap sebelumnya dimuat ke dalam sebuah *data frame* untuk diproses lebih lanjut. Dalam tahapan ini, *data input* diproses dengan menggunakan fungsi khusus yang dirancang untuk menghitung nilai *silhouette score*. Dalam analisis untuk menemukan jumlah *cluster* terbaik, jumlah *cluster* yang dihitung berkisar antara 4 hingga 10. Dengan demikian, beberapa iterasi dilakukan untuk masing-masing nilai *k* dalam rentang ini untuk menentukan *silhouette score* yang paling optimal. Hasil perhitungan *silhouette score* dari algoritma Bisecting K-Means dapat dilihat dalam Gambar 14, yang menunjukkan bagaimana kualitas clustering bervariasi dengan jumlah *cluster* yang berbeda dengan contoh visualisasi *cluster* adalah term *Area* dan term *Data* dimana nilainya adalah jumlah frekuensi *term* yang terdapat pada *dataset*.



Gambar 14 *Silhouette score* dari algoritma Bisecting K-Means

Silhouette score yang dihasilkan oleh algoritma Bisecting K-Means untuk berbagai jumlah *cluster* adalah 0,062; 0,059; 0,054; 0,052; 0,051; 0,067; dan 0,050. Berdasarkan hasil tersebut, dapat disimpulkan bahwa jumlah *cluster* terbaik untuk dataset ini adalah 9 *cluster*, dengan *silhouette score* tertinggi sebesar 0,067. Hal ini menunjukkan bahwa pembagian data ke dalam 9 *cluster* memberikan hasil yang paling optimal, dengan tingkat keseragaman internal *cluster* yang tinggi dan pemisahan antar *cluster* yang jelas. Sebaran *cluster* yang dihasilkan oleh algoritma Bisecting K-Means dipetakan dalam fungsi plot dengan dimensi *data* dan *area*. Hasil pemetaan sebaran *cluster* ini memberikan sebuah gambaran visual tentang pola distribusi data dalam ruang fitur yang diwakili oleh kolom data dan area. Gambar 15 menunjukkan 9 buah *cluster* yang telah terbagi dengan algoritma *regular kmeans* dengan contoh visualisasi *cluster* adalah term *Area* dan term *Data* dimana nilainya adalah jumlah frekuensi *term* yang terdapat pada *dataset*. Anggota-anggota dari tiap *cluster* dapat dilihat pada Lampiran 2.



Gambar 15 Sebaran *cluster* dari algoritma Bisecting K-Means

4.4.3 Clustering dengan algoritma UPGMA

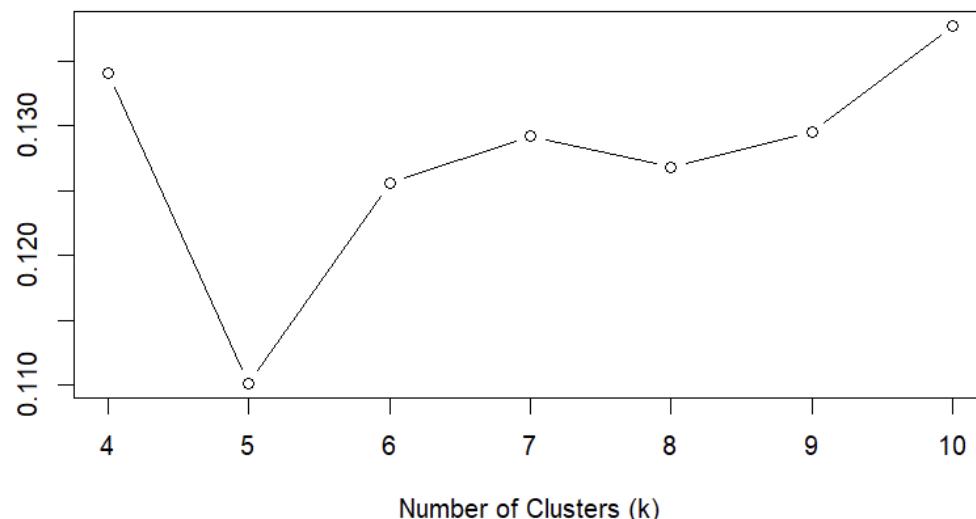
Proses *clustering* dengan algoritma UPGMA dimulai dengan memuat data input yang merupakan data dari *term-term* yang telah diproses sebelumnya ke dalam *dataframe* menggunakan bahasa pemrograman R, yang merupakan langkah awal dalam proses analisis *clustering*. Data sudah yang dimuat didalam *dataframe* kemudian diproses menggunakan berbagai metode untuk *linkage clustering* yang tersedia dalam fungsi *hclust*, seperti contohnya adalah metode *single*, *complete*, *average*, *centroid*, dan *ward*. Setiap metode *clustering* memiliki pendekatan yang berbeda dalam menentukan jarak antar titik data dan pembentukan *cluster*. Penentuan metode *clustering* yang paling sesuai dilakukan berdasarkan evaluasi dari nilai *silhouette score*. Tabel 4 menunjukkan perbedaan nilai *silhouette score* untuk mencari formula linkage terbaik untuk jumlah *cluster* adalah 10 buah *cluster*. Dapat dilihat juga bahwa *silhouette score* yang digunakan pada pencarian metode linkage terbaik menghasilkan nilai yang jauh berbeda dengan nilai *silhouette score* yang lainnya, hal ini dikarenakan terdapat fungsi *cutree* dalam kode untuk memotong dendrogram yang dihasilkan oleh *hclust* sehingga menghasilkan k *cluster* yang kemudian hasil pemotongan dendrogram tersebut dihitung rata-rata dari *silhouette score* sehingga nilai *silhouette score* pada Tabel 4 tampak lebih sempurna daripada hasil *silhouette score* untuk perhitungan lainnya.

Tabel 4 komparasi formula linkage algoritma UPGMA

No	Metode	Silhouette score
1	<i>Single</i>	0,757
2	<i>Complete</i>	0,942
3	<i>Average</i>	0,936
4	<i>Centroid</i>	0,978
5	<i>Ward</i>	1,000

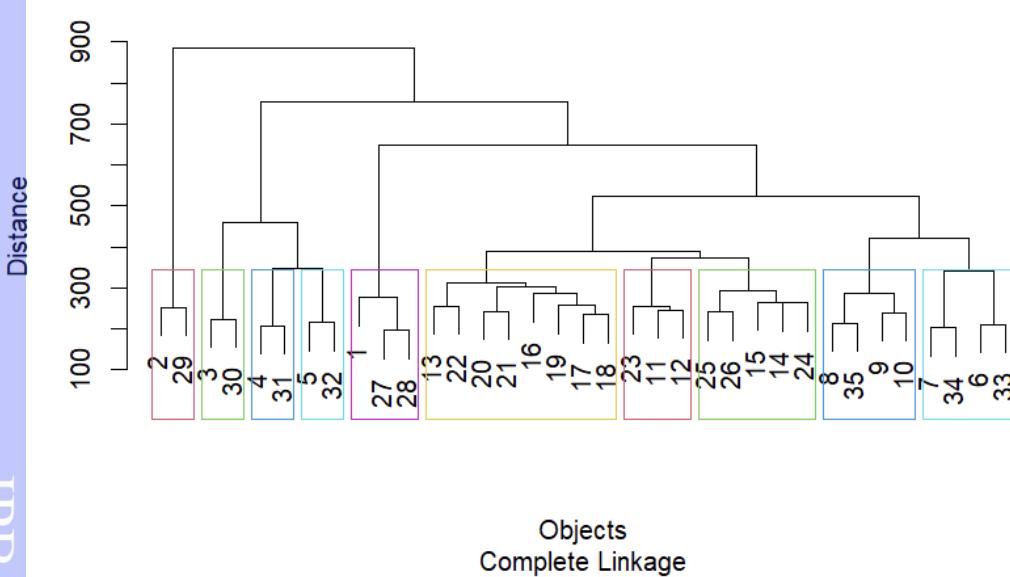
Berdasarkan Tabel 4, metode yang memiliki nilai untuk *linkage* terbaik adalah *linkage ward* dimana nilai *silhouette score* yang dihasilkan adalah 1. Hasil UPGMA *clustering* digunakan sebagai input fungsi *cutree* dengan k buah *cluster*.

Fungsi `cutree` digunakan untuk memotong dendrogram yang dihasilkan dari `hclust` menjadi k buah *cluster*. *Silhouette score* dihitung dengan fungsi `silhouette` dengan input berupa hasil fungsi `cutree`. Perhitungan *silhouette score* dilakukan secara berulang dengan nilai k dari 3 hingga 10. Gambar 16 menunjukkan hasil jumlah *cluster* terbaik yang didapatkan dari algoritma UPGMA *clustering*.



Gambar 16 *silhouette score* dari algoritma UPGMA *Clustering*

Hasil dari algoritma UPGMA *Clustering* memiliki *silhouette score* sebesar 0,134; 0,110; 0,125; 0,129; 0,126; 0,129; dan 0,138. Jumlah *cluster* terbaik berdasarkan *silhouette score* adalah sebanyak 10 buah *cluster* dengan nilai *silhouette score* sebesar 0,138. *Dendrogram* UPGMA *clustering* dipetakan dengan fungsi `cutree`. Gambar 17 menunjukkan pembagian dari 10 *cluster* untuk algoritma UPGMA *clustering* dimana anggota-anggota dari tiap *cluster* dapat dilihat pada Lampiran 2.



Gambar 17 Sebaran *cluster* dari algoritma UPGMA *clustering*

4.5 Evaluasi dan Komparasi Metode

Evaluasi hasil dari setiap metode dilakukan dengan meninjau *silhouette score* yang dihasilkan. *Silhouette score* digunakan untuk mengukur seberapa baik setiap titik data sesuai dengan *cluster* itu sendiri dibandingkan dengan *cluster* lain, yang memberikan indikasi seberapa jelas pemisahan antar *cluster*. Hasil *clustering* dari ketiga metode algoritma *regular K-Means*, *Bisecting K-Means*, dan UPGMA *clustering* memiliki nilai optimum dengan jumlah *cluster* (k) yang berbeda-beda, serta nilai *silhouette score* yang beragam. Komparasi antara ketiga metode dilakukan dengan memperhatikan karakteristik setiap *cluster* yang dihasilkan oleh masing-masing algoritma.

4.5.1 Clustering dengan Algoritma *regular K-Means*

Nilai *silhouette score* terbaik pada algoritma *regular K-Means* adalah sebesar 0,138 dengan jumlah *cluster* sebanyak 10 buah *cluster*. Pada algoritma *regular K-Means*, *cluster* 2, 4, 7, 9, dan 10 lebih berfokus pada *term* atau kata "*area*", sementara *cluster* 1, 3, 5, 6, dan 8 lebih terkait dengan kata "*forest*". Semua *cluster* memiliki keterkaitan erat dengan 6 *term* tertinggi yang ada dikarenakan tingginya nilai *sum* dan nilai *mean* untuk *term-term* yang ada. Karakteristik tiap *cluster regular K-Means* yang menunjukkan nilai mean dan jumlah 6 *term* tertinggi dapat dilihat dalam Tabel 5.

Tabel 5 Karakteristik dari *cluster* algoritma *regular K-Means*

<i>Cluster</i>	Jumlah artikel	Statistik	<i>Terms</i>					
			<i>Area</i>	<i>Forest</i>	<i>Burn</i>	<i>Data</i>	<i>Use</i>	<i>Land</i>
1	3	<i>Median</i>	217	261	178	66	120	172
		<i>Mean</i>	223	281,33	173,33	81,66	121	162,33
		<i>Mode</i>	217	-	-	66	-	-
		<i>Minimum</i>	217	236	159	66	111	139
		<i>Maximum</i>	235	347	184	113	132	176
		<i>Sum</i>	669	844	521	245	363	487
2	7	<i>Median</i>	245	272	166	119	154	158
		<i>Mean</i>	251,28	274,86	170,85	141	152	164,42
		<i>Mode</i>	-	-	-	119	-	-
		<i>Minimum</i>	211	250	158	111	119	136
		<i>Maximum</i>	292	312	197	200	179	198
		<i>Sum</i>	1759	1924	1196	987	1064	1151
3	2	<i>Median</i>	269,5	427,5	203	150,5	208	254
		<i>Mean</i>	269,5	427,5	203	150,5	208	254
		<i>Mode</i>	-	-	-	-	-	-
		<i>Minimum</i>	275	402	180	133	184	236
		<i>Maximum</i>	318	453	226	168	232	272
		<i>Sum</i>	593	855	406	301	416	508
4	4	<i>Median</i>	247,5	243	176	225	210	144
		<i>Mean</i>	251,75	241	176,25	241	209,5	150,5
		<i>Mode</i>	-	-	-	-	-	-
		<i>Minimum</i>	233	201	165	218	179	138

Cluster	Jumlah artikel	Statistik	Terms					
			Area	Forest	Burn	Data	Use	Land
5	4	Maximum	279	277	188	296	239	176
		Sum	1007	964	705	964	838	602
		Median	221	247,5	138	137,5	134,5	150
		Mean	212,5	260,75	138,5	135	135,25	143,75
6	4	Mode	-	-	-	-	-	-
		Minimum	166	221	125	101	124	117
		Maximum	242	327	153	164	148	158
		Sum	850	1043	554	540	541	575
7	2	Median	238	273	197	191	172	159
		Mean	242,4	267,6	197,6	184,6	170,6	161,2
		Mode	-	-	-	-	-	-
		Minimum	220	235	162	171	158	126
8	2	Maximum	278	295	240	196	177	209
		Sum	1212	1338	988	923	853	806
		Median	209,5	201	202	120,5	124,5	134
		Mean	209,5	201	202	120,5	124,5	134
9	2	Mode	-	-	-	-	-	-
		Minimum	203	175	175	118	113	113
		Maximum	216	227	229	123	136	155
		Sum	419	402	404	241	249	268
10	4	Median	204	225	137,5	205,5	173,5	112
		Mean	204	225	137,5	205,5	173,5	112
		Mode	-	-	-	-	-	-
		Minimum	203	199	133	197	170	101
9	2	Maximum	205	211	142	214	177	123
		Sum	408	450	275	411	347	224
		Median	307	245,5	190,5	164	139	165
		Mean	307	245,5	190,5	164	139	165
10	4	Mode	-	-	-	-	-	-
		Minimum	286	235	190	131	136	163
		Maximum	328	256	191	197	142	167
		Sum	614	491	381	328	278	330
10	4	Median	254	233	191,5	104	137,5	163
		Mean	254	241,75	190,75	105,75	135,25	167,25
		Mode	-	-	-	-	-	-
		Minimum	239	213	170	85	111	138
10	4	Maximum	269	288	210	130	155	205
		Sum	1016	967	763	423	541	669

4.5.2 Clustering dengan Algoritma Bisecting K-Means

Nilai *silhouette score* terbaik algoritma *bisecting K-Means* adalah 0,067 dengan jumlah *cluster* sebanyak 9. Pada algoritma *bisecting K-Means*, *cluster* 1 dan 3 memiliki jumlah kata, *mean*, dan *mode* tertinggi pada kata "area". Hal ini menunjukkan bahwa artikel-artikel dalam *cluster* tersebut secara konsisten membahas topik-topik yang berkaitan dengan "area". *Cluster* 2, 4, 5, 6, dan 7 memiliki jumlah kata, *mean*, dan *mode* tertinggi pada kata *forest*. *Cluster* 8

memiliki nilai teringgi pada data *data*. Semua *cluster* berkaitan erat dengan semua kata yang ada berdasarkan jumlah kata dan mean yang tinggi. Karakteristik tiap *cluster bisecting K-Means* yang menunjukkan nilai mean yang tinggi dapat dilihat dalam Tabel 6.

Tabel 6 Karakteristik dari *cluster* algoritma *bisecting K-Means*

<i>Cluster</i>	Jumlah artikel	Statistik	<i>Terms</i>					
			<i>Area</i>	<i>Forest</i>	<i>Burn</i>	<i>Data</i>	<i>Use</i>	<i>Land</i>
1	4	<i>Median</i>	263,4	230,5	178,5	207,5	160,5	155
		<i>Mean</i>	272	229,5	178,25	191,25	168,25	154,5
		<i>Mode</i>	-	-	-	-	-	-
		<i>Minimum</i>	233	201	165	131	136	141
		<i>Maximum</i>	328	256	191	219	218	167
		<i>Sum</i>	1088	918	713	765	675	618
2	5	<i>Median</i>	238	273	197	191	172	159
		<i>Mean</i>	242,4	267,6	197,6	184,6	170,6	161,2
		<i>Mode</i>	-	-	-	-	-	-
		<i>Minimum</i>	220	235	162	171	158	126
		<i>Maximum</i>	278	295	240	196	177	209
		<i>Sum</i>	1212	1338	988	923	853	806
3	2	<i>Median</i>	245	272	166	119	154	158
		<i>Mean</i>	251,25	274,8	170,8	1141	152	164,4
		<i>Mode</i>	-	-	-	119	-	-
		<i>Minimum</i>	211	250	158	111	119	136
		<i>Maximum</i>	292	312	197	200	179	198
		<i>Sum</i>	1759	1924	1196	987	1064	1151
4	2	<i>Median</i>	266,5	268,5	186,5	263,5	220,5	157
		<i>Mean</i>	266,5	268,5	186,5	263,5	220,5	157
		<i>Mode</i>	-	-	-	-	-	-
		<i>Minimum</i>	254	260	185	231	202	138
		<i>Maximum</i>	279	277	188	296	239	176
		<i>Sum</i>	533	537	373	527	441	314
5	6	<i>Median</i>	239	232	172,5	120,5	131,5	155,5
		<i>Mean</i>	228,5	241,2	172,1	120,4	133,1	151,2
		<i>Mode</i>	239	-	-	-	-	-
		<i>Minimum</i>	166	125	85	85	111	113
		<i>Maximum</i>	269	327	229	164	155	205
		<i>Sum</i>	1435	1369	1167	664	790	937
6	3	<i>Median</i>	217	261	178	66	120	172
		<i>Mean</i>	223	281,33	173,33	81,66	121	162,3
		<i>Mode</i>	217	-	-	66	-	3
		<i>Minimum</i>	217	236	236	66	111	-
		<i>Maximum</i>	235	347	347	113	132	139
		<i>Sum</i>	669	844	521	245	363	176

487

Cluster	Jumlah artikel	Statistik	Terms					
			Area	Forest	Burn	Data	Use	Land
7	2	<i>Median</i>	296,5	427,5	203	150,5	208	254
		<i>Mean</i>	296,5	427,5	203	150,5	208	254
		<i>Mode</i>	-	-	-	-	-	-
		<i>Minimum</i>	275	402	180	133	184	236
		<i>Maximum</i>	318	453	226	268	232	272
		<i>Sum</i>	593	855	406	301	416	508
8	4	<i>Median</i>	221	247,5	138	137,5	134,5	150
		<i>Mean</i>	212,5	260,75	138,5	135	135,25	143,7
		<i>Mode</i>	-	-	-	-	-	5
		<i>Minimum</i>	166	221	125	101	124	-
		<i>Maximum</i>	242	327	153	164	148	117
		<i>Sum</i>	850	1043	554	540	541	158
9		<i>Median</i>	204	225	137,5	205,5	173,5	112
		<i>Mean</i>	204	225	137,5	205,5	173,5	112
		<i>Mode</i>	-	-	-	-	-	-
		<i>Minimum</i>	203	199	133	197	170	101
		<i>Maximum</i>	205	251	142	214	177	123
		<i>Sum</i>	408	450	275	411	347	224

4.5.3 Clustering dengan Algoritma UPGMA Clustering

Nilai *silhouette score* terbaik untuk algoritma UPGMA *clustering* adalah 0,138, dengan jumlah *cluster* sebanyak 10. *Cluster* 4, 6, dan 7 memiliki jumlah kata, *mean*, dan *mode* tertinggi pada kata "Area", yang menunjukkan adanya keterkaitan erat dengan topik yang berkaitan dengan area. Sebaliknya, *cluster* 1, 3, 5, 8, dan 10 memiliki jumlah kata, *mean*, dan *mode* tertinggi pada kata "forest". Hal ini mengindikasikan bahwa *cluster-cluster* tersebut memiliki fokus yang lebih kuat pada topik yang berkaitan dengan hutan. Semua *cluster* terkait erat dengan semua kata yang ada yang dapat dilihat berdasarkan jumlah kata dan rata-rata kemunculan yang tinggi. Hal ini menunjukkan adanya keterkaitan dan kemungkinan kesamaan topik atau pendekatan di antara semua *cluster* dalam dataset yang tersedia. Karakteristik dari tiap *cluster* untuk algoritma UPGMA *clustering* yang menunjukkan nilai rataan yang tinggi dapat dilihat dalam Tabel 7.

Tabel 7 Karakteristik dari *cluster* algoritma UPGMA *clustering*

Cluster	Jumlah artikel	Statistik	Terms					
			Area	Forest	Burn	Data	Use	Land
1	1	<i>Median</i>	217	347	159	66	111	176
		<i>Mean</i>	217	347	159	66	111	176
		<i>Mode</i>	-	-	-	-	-	-
		<i>Minimum</i>	217	347	159	66	111	176
		<i>Maximum</i>	217	347	159	66	111	176
		<i>Sum</i>	217	347	159	66	111	176

Cluster	Jumlah artikel	Statistik	Terms					
			Area	Forest	Burn	Data	Use	Land
2	2	<i>Median</i>	296,5	427,5	203	150,5	208	254
		<i>Mean</i>	296,5	427,5	203	150,5	208	254
		<i>Mode</i>	-	-	-	-	-	-
		<i>Minimum</i>	275	402	180	133	184	236
		<i>Maximum</i>	318	453	226	268	232	272
		<i>Sum</i>	593	855	406	301	416	508
3	2	<i>Median</i>	266,5	268,5	186,5	263,5	220,5	157
		<i>Mean</i>	266,5	268,5	186,5	263,5	220,5	157
		<i>Mode</i>	-	-	-	-	-	-
		<i>Minimum</i>	254	260	185	231	202	138
		<i>Maximum</i>	279	277	188	296	239	176
		<i>Sum</i>	533	537	373	527	441	314
4	2	<i>Median</i>	237	213,5	166	218,5	198,5	144
		<i>Mean</i>	237	213,5	166	218,5	198,5	144
		<i>Mode</i>	-	-	-	-	-	-
		<i>Minimum</i>	233	201	165	218	179	141
		<i>Maximum</i>	241	226	167	219	218	147
		<i>Sum</i>	474	427	332	437	397	288
5	2	<i>Median</i>	204	225	137,5	205,5	173,5	112
		<i>Mean</i>	204	225	137,5	205,5	173,5	112
		<i>Mode</i>	-	-	-	-	-	-
		<i>Minimum</i>	203	199	133	197	170	101
		<i>Maximum</i>	205	251	142	214	177	123
		<i>Sum</i>	408	450	275	411	347	224
6	4	<i>Median</i>	277,5	253	190,5	130,5	139	168
		<i>Mean</i>	287	257,25	186,75	135,75	136	176
		<i>Mode</i>	-	-	-	-	-	-
		<i>Minimum</i>	265	235	170	85	111	163
		<i>Maximum</i>	328	288	196	197	155	205
		<i>Sum</i>	1148	1029	747	543	544	704
7	4	<i>Median</i>	227,5	214,5	198,5	111,5	128,5	146,5
		<i>Mean</i>	225,25	207,75	200,25	112,25	131	140,75
		<i>Mode</i>	-	-	-	-	-	-
		<i>Minimum</i>	203	175	175	103	113	113
		<i>Maximum</i>	243	227	229	123	154	157
		<i>Sum</i>	901	831	801	449	524	563
8	3	<i>Median</i>	203	237	142	150	142	144
		<i>Mean</i>	203,66	238,66	140	146,33	138	139,66
		<i>Mode</i>	-	-	-	-	-	-
		<i>Minimum</i>	166	221	125	125	124	117
		<i>Maximum</i>	242	258	153	164	148	158
		<i>Sum</i>	611	716	420	439	414	419
9	8	<i>Median</i>	235	261	178	101	127	156
		<i>Mean</i>	230,33	274,66	165,3	93,33	126,33	155,66
		<i>Mode</i>	-	-	-	-	-	-
		<i>Minimum</i>	217	236	134	66	120	139

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa memantulkan 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 University.

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

Cluster	Jumlah artikel	Statistik	Terms					
			Area	Forest	Burn	Data	Use	Land
10	5	Maximum	239	327	184	113	132	172
		Sum	691	824	496	280	379	467
		Median	242	272	170	170	162	158
		Mean	244,15	269,15	178,92	156,53	157	161,61
		Mode	-	295	197	119	-	-
		Minimum	203	235	142	1111	119	126
		Maximum	292	312	240	200	179	209
		Sum	3174	3499	2326	2035	2041	2101

4.5 Analisis Kriteria dan Indikator Tingkat Keparahan Karhutla

Tahap analisis menggunakan 35 referensi artikel literatur yang dihasilkan pada tahapan *text mining* dan *clustering*. Hasil dari analisis bibliometriks dikaitkan juga dengan hasil dari analisis *clustering*. Bibliometriks menunjukkan bahwa tema-tema penelitian terbanyak adalah *fires*, *Indonesia*, *deforestation*, *burned area*, *remote sensing*, *fire hazard*, *forestry*, *peatland*, *satellite image*, dan *satellite data*. Dari hasil tersebut, dua di antaranya merupakan nilai kriteria tingkat keparahan kebakaran hutan menurut Syaufina (2017), yaitu *burned area* (Wells *et al.* 1979) dan *peatland* (Arstybashev 1983). *Burned area* ditemukan berada pada irisan antara kuadran 1 dan 4, menunjukkan signifikansi dan relevansi tinggi dalam penelitian terkait kebakaran hutan. Hal ini mencerminkan bahwa *burned area* adalah indikator yang sering digunakan untuk menilai tingkat keparahan karhutla di berbagai studi dan juga dalam pengelompokan data yang dihasilkan. *Peatland*, di sisi lain terletak pada kuadran 1, menandakan bahwa topik ini juga sangat relevan dan signifikan dalam penelitian terkait.

Hasil analisis *clustering* yang dilakukan secara manual juga ditemukan bahwa terdapat topik dengan metode atau indikator tingkat keparahan karhutla yang sama seperti yang dikemukakan oleh Syaufina (2017) di dalam hasil *clustering* yang dilakukan dalam penelitian ini, yaitu tanah (*land*), vegetasi (*vegetation*), luas area terbakar (*burned area*), dan penilaian melalui lahan gambut (*peatland*) yang juga muncul sebagai fokus utama dan juga termasuk kedalam *term* terbanyak. Hasil ini menunjukkan adanya korelasi antara temuan dari metode analisis bibliometriks dan analisis *clustering* yang dilakukan. Tabel 8 menunjukkan hasil analisis *cluster* secara manual dimana kriteria dan indikator *burned area* merupakan yang paling banyak digunakan pada tiap algoritma

Hasil *clustering* menunjukkan adanya perbedaan distribusi dari indikator kriteria tingkat keparahan kebakaran hutan dan lahan (karhutla) berdasarkan algoritma-algoritma yang digunakan. Pada algoritma *regular K-Means*, indikator kriteria tingkat keparahan karhutla yang paling banyak muncul adalah indikator *burned area*, dengan sebanyak 6 *cluster* yang mencerminkan bahwa indikator tingkat keparahan ini adalah yang paling banyak digunakan. Selanjutnya adalah indikator *peatland* yang muncul dalam 3 buah *cluster*, dan indikator *vegetation* yang muncul dalam 1 buah *cluster*. Hal ini menunjukkan bahwa *burned area* merupakan indikator yang paling signifikan dalam algoritma *regular K-Means* untuk menentukan tingkat keparahan karhutla.

Pada algoritma *Bisecting K-Means*, pola yang dihasilkan oleh algoritma sedikit berbeda. dari kata *Burned area* tetap menjadi indikator yang paling dominan

atau paling sering muncul, muncul dalam 5 *cluster*, tetapi *peatland* dan *vegetation* masing-masing muncul dalam 2 *cluster*. Ini menandakan bahwa meskipun *burned area* tetap signifikan, ada distribusi yang lebih merata antara *peatland* dan *vegetation* dalam algoritma ini, mencerminkan terdapat pengaruh yang lebih seimbang antara berbagai indikator.

Sementara itu, untuk algoritma UPGMA *clustering*, kriteria tingkat keparahan karhutla terbanyak adalah indikator *burned area* dan *peatland* dimana masing-masing indikator muncul dalam 4 buah *cluster*, sementara indikator *Vegetation* muncul dalam 2 buah *cluster*. Hasil ini menunjukkan bahwa dalam algoritma UPGMA *clustering*, baik indikator *burned area* maupun *peatland* memiliki pengaruh yang hampir setara dalam menentukan tingkat keparahan karhutla, dengan *vegetation* sebagai indikator pelengkap.

Tabel 8 Analisis kriteria dan indikator tingkat keparahan Karhutla

Algoritma	Cluster	Anggota Cluster	Mayoritas kriteria indikator
<i>Regular K-Means</i>	1	3	<i>Burned area</i>
	2	7	<i>peatland</i>
	3	2	<i>Vegetation</i>
	4	4	<i>Burned area</i>
	5	4	<i>Burned area</i>
	6	5	<i>Burned area</i>
	7	2	<i>Burned area</i>
	8	2	<i>Burned area</i>
	9	2	<i>peatland</i>
	10	4	<i>peatland</i>
<i>Bisecting K-Means</i>	1	4	<i>peatland</i>
	2	5	<i>Burned area</i>
	3	7	<i>peatland</i>
	4	2	<i>Vegetation</i>
	5	6	<i>Burned area</i>
	6	3	<i>Burned area</i>
	7	2	<i>vegetation</i>
	8	4	<i>Burned area</i>
	9	2	<i>Burned area</i>
<i>UPGMA Clustering</i>	1	3	<i>Burned area</i>
	2	2	<i>Vegetation</i>
	3	2	<i>vegetation</i>
	4	2	<i>peatland</i>
	5	2	<i>Burned area</i>
	6	4	<i>peatland</i>
	7	4	<i>Burned area</i>
	8	3	<i>peatland</i>
	9	8	<i>peatland</i>
	10	5	<i>Burned area</i>

Selain itu, semua *cluster* terkait erat dengan semua kata yang ada berdasarkan jumlah kata dan *mean* yang tinggi. Hal ini menunjukkan adanya



keterkaitan dan kesamaan topik atau pendekatan di antara semua *cluster* dalam dataset. Pada seluruh algoritma, jumlah artikel dalam setiap *cluster* bervariasi, menunjukkan disparitas dalam representasi topik atau data di antara *cluster-cluster* tersebut. Meskipun terdapat variasi dalam statistik antara *cluster* dan algoritma, beberapa *cluster* menunjukkan konsistensi dalam topik atau tema yang dibahas. Sebagai contoh, terdapat konsistensi dan kesamaan antara *cluster* 1 dalam algoritma UPGMA *clustering* dan *cluster* 1 dalam algoritma *regular K-Means*. Selain itu, terdapat juga kesamaan antara *cluster* 5 dalam algoritma *regular K-Means* dan *cluster* 8 dalam algoritma *bisecting K-Means*. Hasil ini menandakan keberhasilan algoritma pengelompokan dalam mengidentifikasi kelompok artikel yang terkait dengan topik tertentu. Dengan demikian, meskipun terdapat variasi dalam distribusi artikel di antara *cluster*, algoritma-algoritma ini berhasil mengelompokkan artikel berdasarkan kesamaan topik atau pendekatan yang dibahas dalam artikel literatur.

V. SIMPULAN DAN SARAN

5.1 Simpulan

Penelitian ini berhasil mengaitkan hasil analisis bibliometrik yang mengidentifikasi tema-tema penelitian utama dengan temuan dari analisis *clustering*, menegaskan konsistensi dan validitas temuan di antara kedua metode analisis tersebut. Tema analisis bibliometriks yang muncul kedalam indikator dan kriteria tingkat keparahan karhutla adalah *burned area* dan *peatland*, dimana *burned area* terletak beririsan antara kuadran 1 dan 4 dan *peatland* yang terletak pada kuadran 1.

Penelitian ini juga telah berhasil mengelompokkan data teks dengan *clustering* untuk analisis indikator tingkat keparahan karhutla dengan metode *Bisecting K-Means*, *Regular K-Means*, dan *UPGMA Clustering*. Nilai *silhouette score* terbaik dari algoritma *Regular K-Means*, *Bisecting K-Means*, dan *UPGMA clustering* adalah 0,138; 0,067; dan 0,138; dengan jumlah *cluster* sebanyak 10, 9, dan 10. Berdasarkan artikel yang telah dianalisis, kriteria tingkat keparahan kebakaran hutan yang paling sering muncul adalah *burned area* yang diikuti oleh *peatland* dan yang terakhir adalah *vegetation*.

Selain itu, terdapat variasi yang signifikan dalam jumlah artikel, statistik, dan pola antara *cluster-cluster* yang ada. Perbedaan dalam jumlah artikel menunjukkan disparitas dalam representasi topik atau data di antara *cluster-cluster* tersebut yang mencerminkan variasi dalam distribusi data di antara artikel. Pola statistik yang serupa juga terlihat di beberapa *cluster* meskipun algoritma yang dipakai berbeda, menunjukkan adanya kelompok topik yang mirip atau terkait. *Cluster* menunjukkan konsistensi dalam topik atau tema yang dibahas, mengindikasikan keberhasilan algoritma pengelompokan dalam mengidentifikasi kelompok artikel yang berkaitan dengan topik tertentu.

Integrasi temuan dari kedua metode analisis ini mengungkapkan adanya konsistensi dalam fokus penelitian tentang kriteria dan indikator karhutla, dengan indikator yang menjadi fokus utama di kedua hasil analisis, yaitu *burned area* dan *peatland*.

5.2 Saran

Adapun saran yang dapat menjadi acuan untuk penelitian selanjutnya adalah sebagai berikut:

- Analisis artikel literatur yang digunakan masih memiliki keterbatasan pada jumlah artikel sehingga dataset masih kurang besar, dapat digunakan tambahan kombinasi penggunaan kata kunci pada proses pengumpulan data dapat dirincikan atau diperluas agar dapat menemukan referensi artikel literatur yang lebih banyak dan lebih baik
- Mempertimbangkan untuk melakukan algoritma *clustering* dengan jenis algoritma lainnya untuk mencari suatu variasi atau hasil lain dari hasil analisis *clustering*.

DAFTAR PUSTAKA

- Amin A, Ramdani M, Ramdhani A. 2014. Writing a Literature Review Research Paper:A step-by-step approach. *International Journal of Basic and Applied Science*. 3(1): 47-56. doi: <https://digilib.uinsgd.ac.id/5129/>
- Andersen J. 2018. Preaward—Project Preparation. In *The European Research Management Handbook* (pp. 147–171). Elsevier. <https://doi.org/10.1016/B978-0-12-805059-0.00006-7>
- Aria M, Misuraca M, Spano M. 2020. Mapping the Evolution of Social Research and Data Science on 30 Years of Social Indicators Research. *Social Indicators Research*, 149(3): 803–831. <https://doi.org/10.1007/s11205-020-02281-3>
- Baumeister F, Leary R. 1997. Writing narrative literature reviews. *Review of General Psychology*. 1. 311–320. <https://doi.org/10.1037/1089-2680.1.3.311>
- Cronin P, Ryan F, Coughlan M. 2008. Undertaking a literature review: a step-by-step approach. *British Journal of Nursing*. 17(1): 38-43. doi: 10.12968/bjon.2007.16.11.23681
- Damayati A, Utami W, Novitasari D, Kurniawan M. 2023. Cluster Analysis of Environmental Pollution in Indonesia Using Complete Linkage Method with Elbow Optimization. *Jurnal Teori dan Aplikasi Matematika*. 7(2):1-10
- Dang S, Ahmad PH. 2014. Text mining: techniques and its application. *International Journal of Engineering & Technology Innovations*, 1(4), 22-25
- Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. 2021. How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133(1): 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Efendi Z, Sitanggang IS, Syaufina L. 2023. Analisis Dampak Kabut Asap dari Kebakaran Hutan dan Lahan dengan Pendekatan Text Mining. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(5): 1039–1046. <https://doi.org/10.25126/jtiik.20231057248>
- Fadhilah I, Adhi I. 2022. Berapa Luas Negara Indonesia?. [diakses 16 September 2023]. [Internet]. <https://www.kompas.com/global/read/2022/11/01/170000370/berapa-luas-negara-indonesia-?page=all>.
- Fan W, Wallace L, Rich S, Zhang Z. 2006. Tapping the power of text mining. *Communications of the ACM*. 49(9): 76–82.
- Humphreys A, Jen-Hui R. 2018. Automated Text Analysis for Consumer Research, *Journal of Consumer Research*. 44(6):1274–1306, <https://doi.org/10.1093/jcr/ucx104>
- Ivandy F. 2023. Meningkatkan Kewaspadaan Cegah Kebakaran Hutan. [diakses 13 November 2023]. [internet]. <https://ksdae.menlhk.go.id/berita/12016/Meningkatkan-Kewaspadaan-Cegah-Kebakaran-Hutan.html>
- Janani R, Vijayarani S. 2016. Text Mining Research: A Survey. *International Journal of Innovative Research in Computer and Communication Engineering*. 4(4):6564-6571. doi: 10.15680/IJIRCCE.2016.0404040
- Johnson RA, Wichern DW. 1992. *Applied Multivariate Statistical Analysis Third Edition*. New Jersey(US).

- Kambey G, Sengkey R, Jacobus A. 2020. Penerapan Clustering pada Aplikasi Pendekripsi Kemiripan Dokumen Teks Bahasa Indonesia. *Jurnal Teknik Informatika*. 15(2): 339-344
- Kassambara A. 2017. *A practical guide to cluster analysis in R: Unsupervised machine learning*. Marsille(FR):Sthda
- Kaur S, Sohal HS, Cheema RS. 2013. Implementing UPGMA and NJ Method For Phylogenetic Tree Construction Using Hierarchical Clustering. *International Journal Of Computer Science And Technology*. 4(2): 303-310
- Keeley E. 2009. Fire intensity, fire severity and burn severity: a brief review and suggested usage. *International Journal of Wildland Fire*. 18(1): 116-126.
- Khairani NA, Sutoyo E. 2020. Application of K-Means Clustering Algorithm for Determination of Fire-Prone Areas Utilizing Hotspots in West Kalimantan Province. *International Journal of Advances in Data and Information Systems*, 1(1): 9–16. doi: <https://doi.org/10.25008/ijadis.v1i1.13>
- [KLHK] Kementerian Lingkungan Hidup dan Kehutanan. 2023. Nomor: SP. 029 /HUMAS/PPIP/HMS.3/01/2023. Jakarta(ID): Pejabat Pengelola Informasi dan Dokumentasi Kementerian Lingkungan Hidup dan Kehutanan
- Loren A, Ruslan M, Yusran F, Rianawati F. 2015. Analisis Faktor Penyebab Kebakaran Hutan dan Lahan Serta Upaya Pencegahan Yang Dilakukan Masyarakat di Kecamatan Basarang Kabupaten Kapuas Kalimantan Tengah. *EnviroScientiae*. 11(1): 1-9. doi: <http://dx.doi.org/10.20527/es.v11i1.1950>
- Loewenstein Y, Portugaly E, Fromer M, Linial, M. 2008. Efficient algorithms for accurate hierarchical clustering of huge datasets: tackling the entire protein space. *Bioinformatics*. 24(13):41–49. doi: <https://doi.org/10.1093/bioinformatics/btn174>
- Najib M, Nurdiati S, Sopaheluwakan A. 2021. Copula in Wildfire Analysis: A Systematic Literature Review. *Indonesian Journal of Pure and Applied Mathematics*. 3(2):101-111. doi: 10.15408/inprime.v3i2.22131
- Nielsen F. 2016. Hierarchical Clustering. In F. Nielsen, Introduction to HPC with MPI for Data Science. Berlin(DE): Springer
- Notohadinegoro T. 2006. *Pembakaran dan Kebakaran Lahan*. Yogyakarta(ID): Universitas Gadjah Mada Press.
- Oktavia S, Muhlasah N. Neva S. 2013. Pengelompokan Kinerja Dosen JurusanMatematika FMIPA UNTAN BerdasarkanPenilaian Mahasiswa Menggunakan Metode Ward. *Buletin Ilmiah Mat. Stat. dan Terapannya*. 2(2):93-100
- Oti E, Olusola MO, Eze FC, Enogwe SU. 2021. Comprehensive Review of K-Means Clustering Algorithms. *International Journal of Advances in Scientific Research and Engineering*, 7(8): 64–69. doi: <https://doi.org/10.31695/IJASRE.2021.34050>
- Purbowaseso. 2004. *Pengendalian Kebakaran Hutan*. Jakarta(ID): Rineka Cipta.
- Puspitasari N, Widians JA, Setiawan NB. 2020. Customer segmentation using bisecting k-means algorithm based on recency, frequency, and monetary (RFM) model. *Jurnal Teknologi Dan Sistem Komputer*. 8(2): 78–83. doi: <https://doi.org/10.14710/jtsiskom.8.2.2020.78-83>

- Putri I, Nurakhmadyavi S, Wahyudi E. 2022. Literature Review: Sistem Rekomendasi untuk Buku dan Film. *Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika*. 75-81
- Rohilla V, kumar S, Chakraborty S, Singh M. 2019. Data Clustering using Bisecting K-Means. *2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*. [diakses 16 November 2023], doi: 10.1109/ICCCIS48478.2019.8974537
- Samuel E, Adrianto HA, Sitanggang IS. 2023. Analisis Dokumen Publikasi Mengenai Tingkat Emisi Co Dan Co₂ Menggunakan Pendekatan Text Mining. [tesis]. Bogor: Institut Pertanian Bogor
- Shutaywi M, Kachouie N. 2021. Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering. *Entropy*. 23(6): 759. doi: <https://doi.org/10.3390/e23060759>
- Simamora B. 1996. Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique. *J-Climate*. 9(1):130-146
- Singh D, Reddy C. 2014. A survey on platforms for big data analytics. *Journal of Big Data*. 1(8):1-20. doi: 10.1186/s40537-014-0008-6.
- Snyder H. 2019. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*. 104(1): 333–339
- Spurr S, Burton B. 1973. *Forest Ecology*. Minnesota(US):Ronald Press Company
- Syaufina L. 2008. *Kebakaran Hutan dan Lahan di Indonesia, perilaku api, penyebab dan dampak kebakaran*. Bayu Media Publishing, Bandung.
- Syaufina L, Ainuddin N. 2011. Impacts of Fire on SouthEast Asia Tropical Forests Biodiversity: A Review. *Asian Journal of Plant Sciences*. 10(4): 238-244. doi: <https://doi.org/10.3923/ajps.2011.238.244>
- Syaufina L. 2017. Metode Penilaian Areal Pasca Kebakaran Hutan. Bogor(ID): IPB Press.
- Syaufina L, Sitanggang IS, Afina F. 2021. Challenges on fire severity assessment in Indonesia: A vegetation diversity changes perspective. *IOP Conference Series: Earth and Environmental Science*. 886 [diakses 1 November 2023].
- Talib R, Kashif M, Ayesha S, Fatima F. 2016. Text mining: Techniques, applications and issues. *International Journal of Advanced Computer Science and Applications*.7(11):414–418. doi: <https://dx.doi.org/10.14569/IJACSA.2016.071153>
- Templier M, Pare G. 2015. A framework for guiding and evaluating literature reviews. *Communications of the Association for Information Systems*. 37(6):112–137. doi: <https://doi.org/10.17705/1CAIS.03706>
- Webster J, Watson T. 2002. Analyzing the past to prepare for the future: Writing a literature review. *Management Information Systems Quarterly*, 26(2): 13-23. doi: <http://www.jstor.org/stable/4132319>
- Weiß M, Göker M. 2011. *The Yeasts*. San Diego (US):Elsevier. doi: <https://doi.org/10.1016/B978-0-444-52149-1.00012-4>
- Widi S.2023. Luas Kawasan Hutan Indonesia Mencapai 125,76 Juta Hektare. [diakses 16 September 2023]. [internet]. <https://dataindonesia.id/agribisnis-kehutanan/detail/luas-kawasan-hutan-indonesia-mencapai-12576-juta-hektare>



Hak Cipta Dilindungi Undang-undang
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa memantulkan 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 University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

- Yoshida R, Paul L, Nesbitt P. 2022. Stochastic Safety Radius on UPGMA. *Algorithms*. [diakses 2023 Feb 19]; 15(483):2. doi: 10.3390/a15120483.
- Yu J, Muñoz J. 2020. A Bibliometric Overview of Twitter-Related Studies Indexed in Web of Science. *Future Internet*, 12(5): 91. <https://doi.org/10.3390/fi12050091>
- Yudiarta NG, Sudarma M, Ariastina WG. 2018. Penerapan Metode Clustering Text Mining Untuk Pengelompokan Berita Pada Unstructured Textual Data. *Majalah Ilmiah Teknologi Elektro*, 17(3), 339. <https://doi.org/10.24843/MITE.2018.v17i03.P06>
- Zhang C, Huang W, Niu T, Liu Z, Li G, Cao D. 2023. Review of Clustering Technology and Its Application in Coordinating Vehicle Subsystems. *Automotive Innovation*. 6(1): 89-115. doi: <https://doi.org/10.1007/s42154-022-00205-0>
- Zhou Z, Ran A, Chen S, Zhang X, Wei G, Li B, Kang F, Zhou X, Sun, H. 2020. A fast screening framework for second-life batteries based on an improved bisecting K-means algorithm combined with fast pulse test. *Journal of Energy Storage*, 31, 101739. doi: <https://doi.org/10.1016/j.est.2020.101739>.