

ANALISIS DAMPAK KABUT ASAP DARI KEBAKARAN HUTAN DAN LAHAN DENGAN PENDEKATAN *TEXT MINING*

Zuliar Efendi^{*1}, Imas Sukaesih Sitanggang², Lailan Syaufina³

^{1,2,3}Institut Pertanian Bogor, Bogor

Email: ¹zuliarefendizuliar@apps.ipb.ac.id, ²imas.sitanggang@apps.ipb.ac.id, ³lailans@apps.ipb.ac.id

^{*}Penulis Korespondensi

(Naskah masuk: 05 Mei 2023, diterima untuk diterbitkan: 27 September 2023)

Abstrak

Kebakaran hutan dan lahan (karhutla) berdampak buruk bagi lingkungan serta ekosistem. Kabut asap merupakan salah satu akibat yang ditimbulkan dari kebakaran hutan dan lahan. Keresahan dari munculnya kabut asap dan kebakaran hutan menjadi *trending topic* pada media sosial Twitter. Analisis Twitter perlu dilakukan untuk melihat kesesuaian *hashtag* yang digunakan dengan topik yang dibahas yaitu kabut asap. Data Twitter dapat dianalisis menggunakan *text mining*. Penelitian ini bertujuan untuk melihat hubungan antara percakapan di media sosial Twitter dengan kejadian kabut asap yang muncul dari kebakaran hutan dan lahan. Metode yang digunakan adalah teknik *text mining* yaitu menggunakan algoritme *clustering*. Data yang digunakan adalah data *tweet* terkait kabut asap di Provinsi Riau pada jarak 11 – 17 September 2019 dan juga data *hotspot* atau titik panas serta citra Sentinel-2. Data *tweet* dikelompokkan dengan beberapa percobaan pada jarak antar *cluster* yaitu *single linkage*, *complete linkage*, *average linkage*, dan *ward*. Hasil *clustering* menunjukkan bahwa validitas *cluster* tertinggi memiliki *silhouette index* sebesar 0,3360 dengan jarak antar *cluster* menggunakan *ward*. Hasil *cluster* menunjukkan bahwa terdapat tiga *cluster* yang dominan pembahasannya terkait kabut asap. Data Twitter pada ketiga *cluster* tersebut memiliki ciri istilah atau *term* yang berkaitan dengan kabut asap antara lain "kabut", "asap", dan "udara". terdapat di wilayah Pekanbaru serta wilayah Bengkalis, Provinsi Riau. Hasil dapat menjadi salah satu cara pengendalian karhutla yaitu deteksi dini dengan menggunakan media sosial Twitter.

Kata kunci: *Clustering, Kabut Asap, Karhutla, Media Sosial, Text Mining.*

ANALYSIS OF THE IMPACT OF HAZE FROM FOREST AND LAND FIRES WITH A *TEXT MINING APPROACH*

Abstract

Forest and land fires have a harmful impact on the environment and ecosystem. Haze is one of the consequences that arise from forest fires and the environment. Anxiety about haze and forest fires is a trending topic on social media Twitter. Twitter analysis needs to be done to see the compatibility of the hashtags used with the haze topic. The Twitter data can be analyzed using text mining. This study aims to see the relation between conversations on social media Twitter and the occurrence of haze that arises from forest and land fires. The method used is a text mining technique that uses a clustering algorithm. The data used are tweet data related to haze in Riau Province in the range 11-17 September 2019 as well as hotspot data and Sentinel-2 imagery. Tweet data were clustered by several experiments on the distance between clusters, namely single linkage, complete linkage, average linkage, and ward. Clustering results show that the highest cluster validity has a silhouette index of 0.3360 with the distance between clusters using wards. The cluster results show that there are three clusters that are dominant in the discussion related to haze. The Twitter data for the three clusters has the characteristics of terms related to smog, including "kabut", "asap", and "udara". The impact felt by the people of Riau Province through social media Twitter related to the haze is the impact on health and air quality. Cluster tweets that discuss the topic of forest and land fires and haze are in the Pekanbaru and Bengkalis regions, Riau Province. The results can be one of the karhutla controls is early detection by using social media Twitter.

Keywords: *Clustering, Haze, Forest and Land Fires, Social Media, Text Mining*

1. PENDAHULUAN

Hutan merupakan sumber daya alam yang menyediakan berbagai kebutuhan hidup. Menurut

Kementerian Lingkungan Hidup dan Kehutanan (KLHK), Kebakaran hutan dan lahan (Karhutla) adalah peristiwa atau kejadian terbakarnya hutan dan lahan yang dapat disebabkan oleh faktor alami maupun ulah manusia, berakibat pada kerusakan

lingkungan yang menyebabkan dampak ekologi, ekonomi, sosial, budaya, dan politik yang merugikan (KLHK, 2016). Kebakaran hutan dan lahan (Karhutla) bisa terjadi karena disengaja, karena kecelakaan, atau secara alami seperti akibat petir atau suhu panas yang tinggi. Seringkali, kejadian Karhutla tidak dapat dideteksi secara dini. Api dapat dengan cepat menyebar dan menyebabkan kerusakan parah terutama jika terjadi di dekat hutan, daerah pedesaan, wilayah terpencil, dan di sekitar pemukiman penduduk (Yulianti, 2018).

Merujuk dari data SiPongi situs Kementerian Lingkungan Hidup dan Kehutanan, setiap tahun angka titik api terus meningkat yang mengindikasikan dari kebakaran hutan di beberapa wilayah di Indonesia, salah satunya daerah Riau (SiPongi, 2021). Berdasarkan data rekapitulasi kebakaran hutan dan lahan dari SiPongi Kementerian Lingkungan Hidup dan Kehutanan RI hingga 2019 semakin meningkat, salah satu daerah yang terdampak yaitu provinsi Riau dengan luas kebakaran hutan dan lahan hingga 90.233 Ha (KLHK, 2019).

Kebakaran hutan dan lahan memiliki dampak signifikan pada lingkungan, kesehatan manusia, dan perekonomian masyarakat. Banyak dampak yang ditimbulkan oleh kebakaran hutan dan lahan ini banyak sekali seperti yang dirilis oleh worldbank pada Desember 2019 yaitu lebih dari 900.000 orang melaporkan masalah kesehatan pernapasan, 12 bandara nasional dihentikan operasinya, ratusan sekolah di Indonesia serta negara di sekitar lokasi kebakaran hutan harus ditutup sementara, dan kerugian ekonomi pada Juni hingga Oktober 2019 yaitu dengan estimasi sebesar USD 5,2 miliar (Worldbank, 2019). Kejadian karhutla juga membuat munculnya kabut asap dengan skala besar yang dicirikan tingginya konsentrasi bahan partikel, dampak dari kabut asap sendiri dapat bervariasi baik dimulai dengan yang sifatnya lokal seperti menghalangi pandangan hingga yang lebih luas lagi yaitu memungkinkan pemanasan iklim global (Saharjo *et al.*, 2018).

Riau menjadi salah satu daerah kejadian kebakaran hutan dan lahan serta terdampak kabut asap yang dapat dilihat pada Gambar 1 dengan munculnya asap dari karhutla dari salah satu contoh citra Sentinel-2. Hasil *zoom* ke citra pada salah satu daerah Rokan Hilir Provinsi Riau yang dapat dilihat pada Gambar 3 serta didalamnya terdapat penggabungan *band* guna melihat kejadian karhutla. Berdasarkan Gambar 1 dan Gambar 2, menunjukkan terdapat kabut asap yang muncul dari kejadian kebakaran hutan dan lahan di Provinsi Riau.

Kombinasi Band 12, 11, dan 8A pada citra Sentinel-2 (ESA, 2017) ditandai dengan warna coklat-orange yang menjadi area terbakar. Hasil kombinasi Band 12, 11, dan 8A pada Gambar 2 (c) serta *zoom* ke daerah kejadian terbakar dan terdapat kabut asap serta titik panas atau *hotspot*.

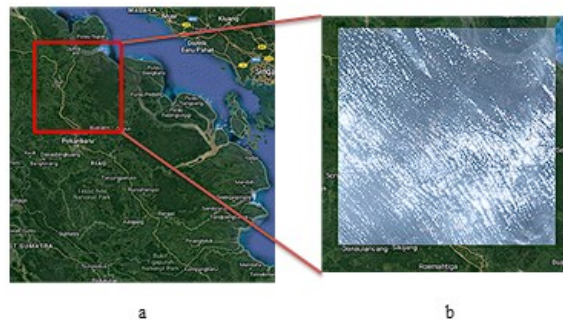
Buruknya kualitas udara di wilayah terdampak kabut asap banyak yang mengalami Infeksi Saluran Pernapasan Akut (ISPA), salah satunya di Pekanbaru Riau sebanyak 9.512 warga yang terserang ISPA (KEMENKES, 2019). Kebakaran hutan dan lahan yang terjadi di beberapa daerah termasuk Provinsi Riau menjadi pembicaraan utama atau *trending topic* yang muncul karena kegelisahan serat keresahan masyarakat dari kebakaran hingga memicu asap yang membahayakan hingga muncul berbagai pembicaraan di media sosial. Salah satu *hashtag* yang banyak digunakan dan menjadi *trending topic* yaitu #RiauDibakarBukanTerbakar (Arumingtyas, 2019). Tweet merupakan salah satu data teks yang merupakan data tidak terstruktur. Sehingga diperlukan sebuah teknik atau metode untuk penanganan data tweet yang menjadi keresahan masyarakat terhadap kabut asap dari karhutla. Percakapan atau *tweet* tersebut dapat dianalisa lebih lanjut dan dimanfaatkan dengan menggunakan *text mining*. *Text mining* mempunyai kemiripan seperti *data mining*, namun bekerja pada *dataset* yang tidak terstruktur atau semi terstruktur seperti data teks (Fan *et al.*, 2006).

Twitter merupakan jaringan sosial jenis layanan *microblogging*. Hal ini memungkinkan pengguna untuk membaca pesan pendek 140 karakter (atau kurang) yang dikenal sebagai *tweet* (Chauhan and Panda, 2015). Twitter digunakan karena memiliki fitur *trending topic* yaitu bahasan yang sedang ramai dibicarakan. *Tweet* menjadi menarik karena terdapat *trending topic* yang menjadi pembicaraan hangat dan banyak dimasyarakat, sehingga perlu ditemukan karakteristik dari *tweet* serta kaitannya dengan konten yang dibicarakan. Salah satu bahasan yang ramai dibicarakan pada Media Sosial Twitter yaitu terkait kabut asap di daerah Riau Tahun 2019.

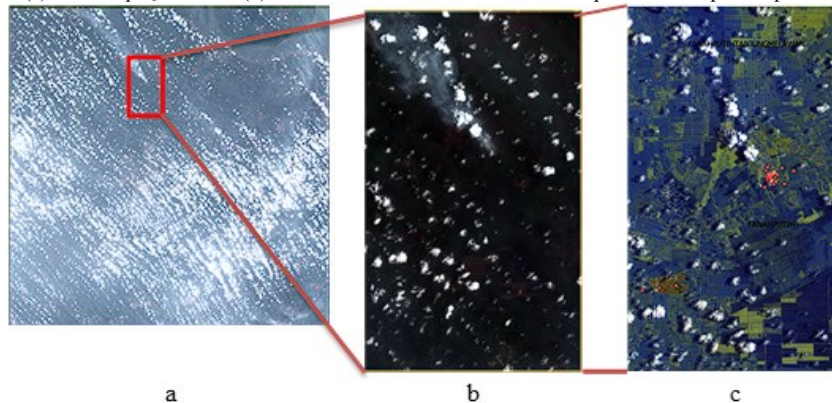
Kabut asap, terutama di daerah Riau menjadi penting karena bukan kejadian pertama kalinya. Sehingga Twitter menjadi salah satu cara untuk memanfaatkan data lainnya selain menggunakan citra satelit, dengan melihat isi atau hal yang ingin disampaikan pada media sosial terkait kabut asap beserta dampak yang dirasakan. Salah satu teknik *text mining* yang mampu untuk melihat sebaran atau pola berdasarkan kemiripan antar objek yaitu *clustering*. Algoritme yang mampu melakukan proses *cluster* secara hirarki dan mampu menampilkan *dendrogram* dalam menentukan banyak jumlah *cluster* yaitu *hierarchical clustering*.

Penelitian sebelumnya terkait kabut asap masih banyak dilakukan menggunakan data citra satelit, untuk data media sosial masih belum banyak dilakukan. Penelitian terkait kabut asap menggunakan media sosial Twitter sudah dilakukan oleh Gurajala *et al.* (2019) namun menggunakan supervised learning pada data tweet pada tiga kota yaitu Paris, London, dan New Delhi dengan kategori yang telah ditentukan diantaranya kesehatan, iklim, politik, atau lainnya (Gurajala, Dhaniyala and

Matthews, 2019). Namun hal tersebut hanya dapat dilakukan pada data yang telah memiliki kelas atau kategori yang telah ditentukan sebelumnya.



Gambar 1. (a) Base map QGIS Riau (b) contoh citra Sentinel-2 daerah terdampak kabut asap 11 September 2019



Gambar 2. (a) Citra 11 September 2019 salah satu daerah di Rokan Hilir (b) Zoom out dari citra a (c) citra b dengan gabungan band yaitu B12, B11, dan B8A serta titik panas atau *hotspot*

Berdasarkan penjelasan di atas, analisis dapat dilakukan terhadap *tweet* tentang dampak kabut asap dari karhutla dengan melihat kecocokan antara kejadian dengan konten *tweet* yang dibicarakan pada media sosial yaitu Twitter. Penelitian ini bertujuan untuk referensi serta informasi penting bagi masyarakat serta evaluasi dari kejadian kabut asap berdasarkan analisis yang dilakukan dari data sosial media yang berkaitan dengan dampak kabut asap dari karhutla dengan pendekatan *text mining*.

2. METODE PENELITIAN

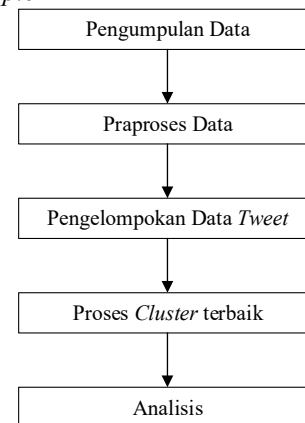
Penelitian analisis dampak kabut asap dari karhutla dengan pendekatan *text mining* memiliki beberapa tahapan, yaitu pengumpulan data, praproses data, pengelompokan data *tweet*, proses menemukan *cluster* terbaik dan analisis seperti pada Gambar 3.

2.1. Pengumpulan Data

Data yang digunakan yaitu citra Sentinel-2, titik panas atau *hotspot*, dan juga *tweet*. Data citra Sentinel-2 dan *hotspot* digunakan untuk melihat lokasi kejadian kebakaran hutan dan lahan seperti pada Gambar 1 dan 2 di daerah yang terdampak kabut asap.

Data *tweet* dikumpulkan dengan menggunakan beberapa kata kunci terkait kejadian karhutla dan kabut asap di Twitter yaitu #IndonesiaDaruratAsap, #RiauDibakarBukanTerbakar, dan

#HukumPembakarHutan dengan menggunakan *library twint* pada Python. Kata kunci tersebut digunakan karena menjadi *trending topic* di media sosial Twitter pada saat kejadian tersebut, diambil sebelum dan sesudah kata kunci tersebut menjadi *trending topic*.



Gambar 3. Tahapan penelitian

Data yang diambil dimulai dari tanggal 11 September 2019 hingga 17 September 2019 di beberapa daerah Kabupaten/Kota di provinsi Riau dengan Indeks Standard Pencemar Udara (ISPU) dengan kategori tidak sehat hingga pada kondisi sangat tidak sehat yang dapat dilihat lebih detail pada Tabel 1. Kondisi tersebut sesuai dengan Peraturan Menteri Lingkungan Hidup dan Kehutanan Republik

Indonesia tentang Indeks Standar Pencemar Udara yaitu Rentang nilai kategori ISPU dari kategori baik hingga sangat berbahaya yang dapat dilihat pada Tabel 2 (KLHK, 2020).

Tabel 1. Nilai Indeks Standar Pencemar Udara (ISPU) Provinsi Riau 14 September 2019 (BNPB, 2019)

No	Kab/Kota	Nilai
1	Pekanbaru	269
2	Dumai	170
3	Rokan Hilir	141
4	Siak	125
5	Bengkalis	121
6	Kampar	113

Tabel 2. Kategori rentang angka ISPU

Kategori	Status Warna	Rentang nilai
Baik	Hijau	1-50
Sedang	Biru	51-100
Tidak Sehat	Kuning	101-200
Sangat Tidak Sehat	Merah	201-300
Berbahaya	Hitam	≥301

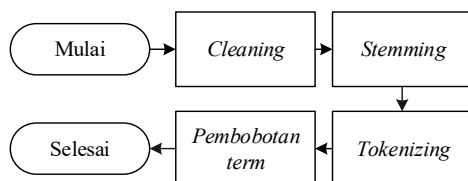
Data *tweet* yang digunakan pada penelitian ini yaitu data *tweet* dengan jumlah sebanyak 646 *tweet* awal. Data tersebut kemudian menjadi 590 setelah dilakukan penghapusan data duplikat yang terlihat pada Tabel 3.

Tabel 3. Data *tweet*

No	<i>Tweet</i>
1	pekanbaru, 11 oktober 2019 #melawanasap
2	semoga asap karhutla selamanya berlalu dari indonesia.. #reformasidikorupsi #melawanasap https://t.co/lrc9fvudi7
3	@liburun #melawanasap terima kasih @ramonytungka 🙏🙏 https://t.co/lqk6wh8rin
4	waspadalah! a great work of @furqonelwe @paksyamsiak @jokowi @wiranto1947 infopku_ info_riau @info_duri beritapekanbaru info_riau walhiriau #melawanasap #masihmelawanasap #riauhaze #haze... https://t.co/wxwthsrneh
5	alhamdulillah, pekanbaru pagi ini #melawanasap #riau https://t.co/pdr2i4sh5i
...	...
590	bpjs tekor triliunan mau minta tolong china yang lebih jauh? #indonesiadaruratasap malah ngeles "menjaga harkat dan martabat negara" dari tetangga terdekat yang ikut kena dampak bencana?? #nalar 🤔

2.2. Praproses Data

Praproses dilakukan untuk mendapatkan data yang siap dilakukan proses penerapan algoritme *clustering*. Praproses mempunyai beberapa tahap diantaranya *cleaning*, *stemming*, *tokenizing*, serta pembobotan seperti yang terlihat pada alur praproses data pada Gambar 4.



Gambar 4. Tahapan praproses data

Tahapan *cleaning* merupakan proses menghapus angka, simbol, karakter khusus, tanda

baca, *link* serta *stopword* atau kata muncul namun dianggap tidak memiliki arti. *Stoplist* yang digunakan pada penelitian ini berjumlah 758 kata dari penelitian Tala (Tala, 2003) terlihat pada Tabel 3. Proses *stemming* adalah suatu metode untuk mengubah berbagai variasi morfologi kata atau kalimat menjadi bentuk dasar yang seragam (Siswandi, Permana and Emarilis, 2021). *stemming* dilakukan menggunakan algoritme Nazief dan Adriani (Adriani *et al.*, 2007). *Tokenizing* merupakan langkah dari pra proses guna menghasilkan token dari masing-masing dokumen (Panda, 2018). Praproses terakhir adalah pembobotan pada setiap kata dari *tweet* berdasarkan tingkat frekuensinya.

Tahapan terakhir yaitu pembobotan, dalam proses formalisasi dokumen, dokumen diwakili oleh vektor dokumen yang diharapkan menunjukkan sebanyak mungkin informasi dari dokumen tersebut. Istilah atau *term* bobot mempunyai peran penting, karna merepresentasikan isi dokumen (Mehare and Deorankar, 2018).

2.3. Pengelompokkan Data *Tweet*

Tahap pengelompokkan, *dataset* yang dilakukan adalah proses *cluster/* pengelompokkan dari konten *tweet* pengguna Twitter dan *tweet* yang terkait dengan kejadian karhutla serta dampaknya. Proses *clustering* dilakukan dengan menggunakan *hierarchical* yaitu pengelompokkan yang dimulai pada setiap objek itu sendiri artinya banyak *cluster* sama dengan banyak objek. Objek-objek yang memiliki kesamaan akan digabungkan menjadi satu kelompok dan seterusnya hingga membentuk satu kelompok utuh atau *cluster* (Johnson and Wichieni, 2007).

Salah satu algoritme hirarki yaitu *Agglomerative clustering* yang merupakan pengelompokkan hirarki yang dimulai dari objek dalam *cluster* terpisah kemudian bergabung menjadi *cluster* yang besar berdasarkan hitungan jarak (Vijaya, Aayushi and Bateja, 2017). Proses pengelompokkan hirarki adalah sebagai berikut (Vijaya, Sharma and Batra, 2019).

1. Tahap pertama, setiap *item* diperlakukan sebagai *cluster* independen.
2. Gunakan jarak sebagai sebagai metrik untuk mengukur kesamaan, kemudian temukan dua pasangan dari *item* terdekat dan kemudian menjadi satu *cluster*.
3. Kemudian hitung lagi jarak antara *cluster* yang baru dibuat dengan *cluster* lama.
4. Lakukan kembali proses 2 dan 3 secara berulang hingga hanya ada satu kelompok yang berisi semua *item* atau data.

Tahap 2 atau 3, percobaan dilakukan dengan beberapa jarak antar *cluster* yaitu *single linkage*, *average linkage*, *complete linkage*, dan juga *ward*. *Single linkage* yaitu mendefinisikan jarak antara dua *cluster* berdasarkan jarak minimum setiap pasangan

anggota dari dua *cluster* (Li, Rezaeipanah and Tag El Din, 2022). *Average linkage* adalah metode yang menentukan jarak antara dua *cluster* berdasarkan jarak rata-rata dari semua anggota yang ada dalam *cluster* tersebut.. *Complete linkage* merupakan metode yang menentukan jarak antara dua *cluster* berdasarkan jarak maksimum antara semua anggota dalam *cluster*. Terakhir yaitu *ward*, metode yang berdasarkan *Sum of Square Error* (SSE). Setiap kemungkinan penyatuan *cluster* dipertimbangkan, yang dalam penggabungannya menghasilkan nilai minimum dalam kehilangan informasi (Vijaya, Sharma and Batra, 2019).

Kemudian dilanjutkan dengan penentuan jumlah *cluster* berdasarkan *dendrogram* yang didapat hingga menemukan *cluster* optimal. *Dendrogram* menunjukkan struktur *tree* dari jaringan hierarki mulai dari *root node* yang kemudian dipecah secara berurutan hingga *leaf node* (Vinagre Díaz *et al.*, 2020).

2.4. Evaluasi Cluster

Cluster optimal ditentukan dengan validitas *cluster silhouette index* dengan melihat tingkat baiknya sebuah objek jika diletakkan pada suatu *cluster*. Validasi *cluster* ini berguna untuk menentukan kualitas *cluster* dari setiap percobaan yang kemudian diambil dari hasil optimal yang kemudian akan dilakukan analisa lebih lanjut. Evaluasi *cluster* dilakukan menggunakan *silhouette index*. *Silhouette Index* merupakan sebuah metode untuk menentukan kualitas dan kekuatan dari *cluster* dengan cara melihat tingkat baik serta bagus nya sebuah data atau objek jika diletakkan pada suatu *cluster*. Rata-rata dari *silhouette* menunjukkan evaluasi dari validitas pengelompokan yang digunakan untuk memilih atau menemukan jumlah *cluster* yang sesuai (Rousseeuw, 1987). Adapun persamaan untuk *silhouette* dapat dilihat pada Persamaan (1).

$$S(i) = \frac{b(i)+a(i)}{\max\{a(i),b(i)\}} \quad (1)$$

dengan:

$b(i)$ = rata-rata jarak objek ke- i pada setiap objek pada satu kelompok.

$a(i)$ = rata-rata jarak objek ke- i pada setiap objek pada kelompok yang berbeda.

$s(i)$ = nilai koefisien *silhouette*.

2.5. Analisis Hasil Clustering

Analisis dilakukan pada *cluster* optimal dengan melihat tingkat frekuensi munculnya kata atau *term* pada setiap *cluster*. Tahap ini juga melihat karakteristik dari setiap *cluster* yang kemudian dicocokkan dengan dampak kabut asap terhadap konten yang dibahas.

3. HASIL DAN PEMBAHASAN

Hasil dan pembahasan dimulai dari praproses data hingga ke tahap analisis. Detail hasil serta pembahasan pada penelitian ini adalah sebagai berikut.

3.1. Hasil Praproses Data

Praproses terdiri 4 tahap seperti yang sudah dijelaskan yaitu *cleaning*, *stemming*, *tokenizing* dan pembobotan *term*. Tahap *cleaning* hasil *tweet* awal dan setelahnya dilakukan proses penghapusan duplikat *tweet* sehingga menjadi 494 *tweet* dari 590 *tweet* dan *stopword* terlihat pada Tabel 4.

Tabel 4. Contoh hasil penghapusan *stopword*

No	<i>Tweet</i>	Hasil penghapusan <i>stopword</i>
1	pekanbaru oktober	pekanbaru oktober
2	semoga asap karhutla selamanya berlalu dari indonesia	semoga asap karhutla indonesia
3	terima kasih	terima kasih
4	waspadalah a great work of infopku info riau duri beritapekanbaru info riau walhiriau	waspadalah a great work of infopku info riau duri beritapekanbaru info riau walhiriau
5	alhamdulillah pekanbaru pagi ini	alhamdulillah pekanbaru pagi
...
494	bpjs tekor triliunan mau minta tolong china yang lebih jauh malah ngeles menjaga harkat dan martabat negara dari tetangga terdekat yang ikut kena dampak bencana	bpjs tekor triliunan tolong china ngeles menjaga harkat martabat negara tetangga terdekat kena dampak bencana

Selanjutnya yaitu tahap atau proses mengembalikan kata berimbuhan menjadi kata dasar atau *stemming*. Contoh hasil *stemming* terlihat pada Tabel 5.

Tabel 5. Contoh hasil *stemming*

No	<i>Tweet</i>	Hasil <i>stemming</i>
1	pekanbaru oktober	pekanbaru oktober
2	semoga asap karhutla selamanya berlalu dari indonesia	moga asap karhutla indonesia
3	terima kasih	terima kasih
4	waspadalah a great work of infopku info riau duri beritapekanbaru info riau walhiriau	waspada a great work of infopku info riau duri beritapekanbaru info riau walhiriau
5	alhamdulillah pekanbaru pagi ini	alhamdulillah pekanbaru pagi
...
494	bpjs tekor triliunan mau minta tolong china yang lebih jauh malah ngeles menjaga harkat dan martabat negara dari tetangga terdekat yang ikut kena dampak bencana	bpjs tekor triliun tolong china ngeles jaga harkat martabat negara tetangga dekat kena dampak bencana

terbanyak yaitu Pekanbaru dan Bengkalis. *Wordcloud* karakteristik *cluster* 2 dapat dilihat pada Gambar 8.



Gambar 8. *Wordcloud cluster 2*

Cluster 3 lebih mengarah pada konten terkait kebakaran hutan dan lahan yang didalamnya ada keterkaitannya dengan kabut asap. Namun tidak dengan frekuensi yang banyak seperti pada *cluster-cluster* lainnya karena merupakan dampak dari kejadian karhutla, daerah yang mendominasi pada *cluster* Pekanbaru dan juga Bengkalis. *Wordcloud* karakteristik *cluster* 3 dapat dilihat pada Gambar 9.



Gambar 9. *Wordcloud cluster 3*

Terakhir pada *cluster* 4 tidak terlalu banyak *term* yang berkaitan langsung dengan kejadian kabut asap muncul pada *cluster* ini. Namun tetap masih muncul terkait kabut asap dan dampaknya seperti *term* "ispa" yang sempat ikut muncul, namun sedikit sekali dibandingkan dengan sebelumnya dan didominasi daerah Pekanbaru. *Wordcloud* karakteristik *cluster* 4 dapat dilihat pada Gambar 10.



Gambar 10. *Wordcloud cluster 4*

Berdasarkan beberapa analisis pada setiap *cluster* yang telah dijelaskan. Setiap *cluster* memuat pembicaraan terkait kabut asap. Dampak yang dirasakan oleh masyarakat berdasarkan *tweet trending topic* terkait keresahan dan kegelisahan kabut asap dari karhutla pada media sosial Twitter adalah masalah kesehatan serta kualitas udara terlihat banyak *term* yang muncul terkait kedua hal tersebut.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil yang didapatkan pada penelitian ini, *cluster* optimal ditemukan pada jumlah *cluster* 4 menggunakan jarak antar *cluster ward*, dengan nilai SI 0,3339. Karakteristik dari ke-4 *cluster*

memiliki persamaan dari *term* yang muncul. Terlihat dari frekuensi terkait kabut asap yang mendominasi. Namun, pada *cluster* 3 juga memiliki karakteristik yang berkaitan dengan kebakaran hutan dan lahan, kemudian diikuti juga dengan *term* terkait kabut asap namun tidak dominan dibandingkan yang lainnya. Kelompok tersebut dapat digunakan untuk menentukan tingkat antisipasi, informasi dan peringatan terhadap kabut asap.

Hasil *cluster* yang dihasilkan pada percobaan didapatkan bahwa dampak yang diterima masyarakat berdasarkan media sosial Twitter yaitu masalah kesehatan serta kualitas udara yang terlihat dari *term* yang muncul pada setiap hasil *cluster*. Sehingga penelitian ini dapat menjadi salah satu cara dalam pengendalian karhutla yaitu deteksi dini selain dengan metode yang telah digunakan saat ini oleh instansi terkait tentang indikasi karhutla.

Dalam penelitian selanjutnya, percobaan dapat dilakukan dengan menambah data set dari beberapa tahun kejadian kabut asap sebelumnya dan di daerah yang berbeda. Hasil dari kelompok yang didapat pada data *tweet* bisa dibuat lebih berguna lagi misalnya saat data baru yang masuk yaitu berupa *tweet*, percakapan itu dapat langsung identifikasi dengan klasifikasi dan dilakukan tindak lanjut.

DAFTAR PUSTAKA

- ADRIANI, M., ASIAN, J., NAZIEF, B., TAHAGHOGHI, S. AND WILLIAMS, H.E., 2007. Stemming Indonesian: A confix-stripping approach. *Stemming Indonesian: A Confix-Stripping Approach. ACM Transactions on Asian Language Information Processing*, 6(4), pp.13–33. <https://doi.org/10.1145/1316457.1316459>.
- ARUMINGTYAS, L., 2019. *Bencana Asap di Sumatera dan Kalimantan, Mengapa Lahan Gambut Terus Terbakar?* [online] Mongabay. Available at: <<https://www.mongabay.co.id/2019/09/15/bencana-asap-di-sumatera-dan-kalimantan-mengapa-lahan-gambut-terus-terbakar/>> [Accessed 8 December 2022].
- BNPB, 2019. *Kualitas Udara Riau Masih Buruk*. [online] BNPB. Available at: <<https://bnpb.go.id/berita/kualitas-udara-riau-masih-buruk>> [Accessed 7 December 2022].
- CHAUHAN, S. AND PANDA, N., 2015. Open Source Intelligence and Advanced Social Media Search. pp.15–32. <https://doi.org/10.1016/B978-0-12-801867-5.00002-1>.
- ESA, 2017. *Training Kit - HAZA02 Burned Area Mapping With Sentinel-2 using Snap*.
- FAN, W., WALLACE, L., RICH, S. AND ZHANG, Z., 2006. Tapping the power of text mining. *Communications of the ACM*, 49(9), pp.76–

82.
<https://doi.org/10.1145/1151030.1151032>.
- GURAJALA, S., DHANIYALA, S. AND MATTHEWS, J.N., 2019. Understanding Public Response to Air Quality Using Tweet Analysis. *Social Media and Society*, 5(3). <https://doi.org/10.1177/2056305119867656>.
- JOHNSON, R.A. AND WICHIERN, D.W., 2007. Applied Multivariate Statistical Analysis. Pearson Education, New Jersey: Pearson Prentice Hall.
- KEMENKES, 2019. Udara di Riau Capai Level Berbahaya. [online] KEMENKES. Available at: <<https://www.kemkes.go.id/article/view/19091600003/udara-di-riau-capai-level-berbahaya.html>> [Accessed 8 December 2022].
- KLHK, 2016. Peraturan Menteri Lingkungan Hidup dan Kehutanan No P.32/MenLHK/Setjen/Kum.1/3/2016. KLHK.
- KLHK, 2019. Rekapitulasi Luas Kebakaran Hutan dan Lahan (Ha) Per Provinsi di Indonesia Tahun 2014-2019. [online] Available at: <http://sipongi.menlhk.go.id/pdf/luas_kebakaran> [Accessed 20 February 2020].
- KLHK, 2020. PERATURAN MENTERI LINGKUNGAN HIDUP DAN KEHUTANAN REPUBLIK INDONESIA NOMOR P.14/MENLHK/SETJEN/KUM.1/7/2020 TENTANG INDEKS STANDAR PENCEMAR UDARA.
- LI, T., REZAEIPANAH, A. AND TAG EL DIN, E.M., 2022. An ensemble agglomerative hierarchical clustering algorithm based on clusters clustering technique and the novel similarity measurement. *Journal of King Saud University - Computer and Information Sciences*, 34(6, Part B), pp.3828–3842.
<https://doi.org/https://doi.org/10.1016/j.jksuci.2022.04.010>.
- MEHARE, D.D. AND DEORANKAR, A. V., 2018. Introduction to TF-IDF: To Represent Importance of Keyword within whole Dataset. *International Journal for Research in Applied Science and Engineering Technology*, 6, pp.2321–2323.
- PANDA, M., 2018. Developing an Efficient Text Pre-Processing Method with Sparse Generative Naive Bayes for Text Mining. *I.J. Modern Education and Computer Science*, 9, pp.11–19.
<https://doi.org/10.5815/ijmecs.2018.09.02>.
- ROUSSEEUW, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C), pp.53–65.
[https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- SAHARJO, B.H., SYAUFINA, L., NURHAYATI, A.D., PUTRA, E.I., WALDI, R.D. AND WARDANA, 2018. Pengendalian Kebakaran Hutan dan Lahan di Wilayah Komunitas Terdampak Asap. IPB Press. PT Penerbit IPB Press.
- SIPONGI, 2021. Luas Karhutla. [online] KLHK2. Available at: <<https://sipongi.menlhk.go.id/>> [Accessed 20 February 2020].
- SISWANDI, A., PERMANA, A.Y. AND EMARILIS, A., 2021. Stemming Analysis Indonesian Language News Text with Porter Algorithm. *Journal of Physics: Conference Series*, 1845, pp.1–7.
<https://doi.org/10.1088/1742-6596/1845/1/012019>.
- TALA, F.Z., 2003. A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia. *M.Sc. Thesis, Appendix D*, pp. pp.39–46.
- VIJAYA, AAYUSHI, S. AND BATEJA, R., 2017. A Review on Hierarchical Clustering Algorithms. *Journal of Engineering and Applied Sciences*, 12(24), pp.7501–7507.
- VIJAYA, SHARMA, S. AND BATRA, N., 2019. Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering. In: *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*. pp.568–573.
<https://doi.org/10.1109/COMITCon.2019.8862232>.
- VINAGRE DÍAZ, J.J., FERNÁNDEZ POZO, R., RODRÍGUEZ GONZÁLEZ, A.B., WILBY, M.R. AND SÁNCHEZ ÁVILA, C., 2020. Hierarchical Agglomerative Clustering of Bicycle Sharing Stations Based on Ultra-Light Edge Computing. *Sensors*, 20(12).
<https://doi.org/10.3390/s20123550>.
- WORLD BANK, 2019. Indonesia Economic Quarterly Investing in People December 2019. [online] Available at: <<https://openknowledge.worldbank.org/bitstream/handle/10986/33033/Investing-in-People.pdf?sequence=1&isAllowed=y>> [Accessed 23 March 2020].
- YULIANTI, N., 2018. Pengenalan Bencana Kebakaran dan Kabut Asap Lintas Batas [Studi Kasus Eks Proyek Lahan Gambut Sejuta Hektar]. PT Penerbit IPB Press.