

PERBANDINGAN METODE CLUSTERING MENGGUNAKAN HIERARCHICAL CLUSTERING DAN PARTITIONAL CLUSTERING UNTUK MENGELOMPOKKAN DOKUMEN BERITA

Yunita Kusumawardani¹, Amir Hamzah², Suraya³

^{1,2,3} Teknik Informatika, Institut Sains & Teknologi AKPRIND, Yogyakarta
yunitakusumawardani@yahoo.com, amir@akprind.ac.id, suraya@akprind.ac.id

ABSTRACT

The development of the internet in Indonesia increasingly advanced and sophisticated. Similarly mass media that has been a lot of venturing into the media online to facilitate users in accessing the news or information via online. And here began of growing and abundant news or information in the text. So make the news has not categorized properly. Then cause difficulties in searching appropriate news categories and should spend a lot of time in searching the abundant news text on the internet.

Research to classsifying this news of the text, using 2 clustering method that is Hierarchical Clustering by using Agglomerative Hierarchical Clustering and Partitional Clustering method using K-Means Clustering. Two clustering methods are used, with the aim of knowing and comparing which methods are better and resulting in higher accuracy in grouping of news texts, as well as faster time in grouping processing.

This research use data totaling 30 news data. From 30 news data is divided into 8 news data for news document collection 1, 10 news data for news document collection 2, and 12 news data for news document collection 3. The result of research using agglomerative hierarchical clustering obtained accuracy value on news document collection 1 that is 62,5%, news document collection 2 is 50%, news document collection 3 is 33,33%, for k-means clustering is obtained result of highest accuracy value in news document collection 1 is 62,5%, news document collection 2 is 60 %, News document collection 3 is 33,33%. While the average time obtained with agglomerative hierarchical clustering is 29,5 s and k-means clustering is 30,19 s.

Keywords: *News of the text, K-means Clustering, Agglomerative Hierarchical Clustering, accuracy*

INTISARI

Perkembangan internet di Indonesia semakin bertambah maju dan canggih. Sama halnya dengan media massa yang sudah banyak merambah ke media online untuk memudahkan pengguna mengakses berita atau informasi secara *online*. Dan disinilah mulai bertambah besar dan melimpah berita atau informasi berupa teks. Sehingga menjadikan berita belum terkategori atau dikelompokkan dengan baik. Kemudian menyebabkan kesulitan dalam pencarian berita yang tepat sesuai kategori dan harus meluangkan banyak waktu dalam menelusuri teks berita yang melimpah di internet.

Penelitian untuk mengelompokkan berita teks ini, menggunakan 2 metode *clustering* yaitu *Hierarchical Clustering* dengan menggunakan metode *Agglomerative Hierarchical Clustering* dan *Partitional Clustering* menggunakan metode *K-Means Clustering*. Digunakan 2 metode *clustering*, dengan tujuan untuk mengetahui serta membandingkan metode mana yang lebih bagus dan menghasilkan akurasi yang lebih tinggi dalam pengelompokkan teks berita, serta waktu yang lebih cepat dalam pemrosesan pengelompokkan.

Penelitian ini menggunakan data yang berjumlah 30 data berita. Dari 30 data berita dibagi menjadi 8 data berita untuk koleksi dokumen berita 1, 10 data berita untuk koleksi dokumen berita 2, dan 12 data berita untuk koleksi dokumen berita 3. Hasil penelitian menggunakan *agglomerative hierarchical clustering* diperoleh nilai akurasi pada koleksi dokumen berita 1 yaitu 62,5%, koleksi dokumen berita 2 yaitu 50%, koleksi dokumen berita 3 yaitu 33,33%, untuk k-

means clustering diperoleh hasil nilai akurasi paling tinggi pada koleksi dokumen berita 1 yaitu 62,5%, koleksi dokumen berita 2 yaitu 60%, koleksi dokumen berita 3 yaitu 33,33%. Sedangkan rata-rata waktu yang diperoleh dengan *agglomerative hierarchical clustering* yaitu 29,5 s dan *k-means clustering* yaitu 30,19 s.

Kata kunci: berita teks, *K-means Clustering*, *Agglomerative Hierarchical Clustering*, akurasi

PENDAHULUAN

Dari tahun ketahun perkembangan internet di Indonesia semakin bertambah maju dan canggih, dan juga pengguna internet di Indonesia semakin bertambah banyak. Sama halnya dengan media massa yang sudah banyak merambah ke media online untuk memudahkan pengguna mengakses berita atau informasi secara *online*.

Media online (*online media*) adalah media massa yang tersaji secara *online* di situs web (website) internet (Romli, 2012). Masih menurut Romli (2012) dalam buku tersebut, media *online* adalah media massa “generasi ketiga” setelah media cetak (*printed media*), seperti: koran, tabloid, majalah, buku, dan media elektronik (*electronic media*), seperti: radio, televisi, dan film/video.

Hal ini dapat dilihat dengan bermunculannya situs-situs berita online antara lain: detik.com, kompas.com, liputan6.com, vivanews.com, okezone.com, tempo.co, Jawapos.com, dan lain sebagainya. Dan disinilah mulai bertambah besar dan melimpah berita atau informasi berupa teks. Oleh sebab itu sangat dibutuhkan metode untuk mengelompokkan berita-berita berupa teks tersebut agar teks berita dapat tekategori dengan baik. Metode tersebut adalah metode *clustering*. *Clustering* adalah metode penganalisaan data, yang sering dimasukkan sebagai salah satu metode dalam *data mining*. Tujuannya untuk mengelompokkan data. Di dalam metode *clustering* ini terdapat dua pendekatan yaitu *Hirearchical* (hirarki) dan *Partitioning* (partisi), maka dari itu pengelompokkan berita akan dikelompokkan menggunakan *Hierarchical Clustering* dan *Partitional Clustering*. *Clustering* dengan pendekatan hirarki atau *hierarchical clustering* terlebih dahulu dilakukan pengelompokkan terhadap dua atau lebih objek yang memiliki karakteristik atau kemiripan paling dekat. Sedangkan *clustering* dengan pendekatan non hirarki atau partisi atau *partitional clustering* menentukan terlebih dahulu jumlah kelompok atau *cluster* yang diinginkan. Setelah itu, proses *clustering* dilakukan untuk mengetahui karakteristik (kemiripan) yang ada pada masing-masing *cluster* (Hidayat dan Istiadah, 2011). Metode hirarki yang digunakan adalah *Agglomerative Hierarchical Clustering* (AHC). Dan metode partisi yang sering digunakan adalah *k-means*. Menurut Atmajaya (2016), kelebihan *k-mean* antara lain: (1) Mudah untuk diimplementasikan dan dijalankan, (2) Waktu yang dibutuhkan untuk menjalankan pembelajaran ini relatif cepat, (3) Mudah untuk diadaptasi, dan (4) Umum digunakan. Sedangkan kekurangan *k-means* yaitu sebelum algoritma dijalankan, *k* buah titik diinisialisasi secara *random* sehingga pengelompokkan data yang dihasilkan dapat berbeda-beda. Jika nilai *random* untuk inisialisasi kurang baik, maka pengelompokkan yang dihasilkan pun menjadi kurang optimal.

TINJAUAN PUSTAKA

Dalam penelitian ini menggunakan beberapa referensi yang berkaitan dengan obyek dan kasus penelitian yang dilakukan. Beberapa referensi diambil dari sumber yang berhubungan dengan metode AHC dan *metode K-means Clustering*, serta hasil-hasil penelitian yang telah dilakukan sebelumnya.

Penelitian yang dilakukan oleh Kusumawardani di tahun 2016 yang berkaitan dengan pengelompokan opini, menjadi opini negatif dan opini positif yang diambil dari opini mahasiswa yang dikelompokkan dengan algoritma *k-means*, dimana pengujian menggunakan 4 koleksi dokumen yang masing-masing dokumen terdapat jumlah opini yang berbeda-beda. Koleksi dokumen 1 berisi 10 opini, koleksi dokumen 2 berisi 30 opini, koleksi dokumen 3 berisi 40 opini, dan koleksi dokumen 4 berisi 50 opini. Diperoleh hasil akurasi tertinggi 90% dan akurasi rata-rata keseluruhan dokumen yang diujikan diperoleh 52,48%. Pada penelitian ini data dokumen yang diujikan perkalimat masih sedikit (Kusumawardani, 2016).

Penelitian yang dilakukan oleh Husni dkk ditahun 2015 yang berkaitan dengan pengelompokan dokumen web (berita) dari salah satu situs perguruan tinggi di Indonesia yaitu Universitas Trunojoyo Madura yang ditulis dengan bahasa Indonesia, dimana pengujiannya menggunakan acuan data berupa sekumpulan dokumen yang berjumlah 300 dokumen berita yang telah diklasifikasi atau dikelompokkan secara manual berdasarkan kemiripan berita. Kemudian hasil *cluster* dengan *K-means* dihitung nilai *F-Measure* dan *Purity* berdasarkan kelas-kelas dokumen berdasarkan pengelompokan secara manual oleh 3 orang berbeda. Diperoleh hasil berdasarkan perhitungan dari beberapa percobaan yaitu perbandingan nilai *F-measure* yang paling besar adalah percobaan dengan jumlah *cluster* = 2 dengan nilai *F-Measure* untuk *cluster* = 2 sebesar 0.730666369, dengan nilai rata-rata 0.612919. Dan nilai *Purity* paling besar pada jumlah *cluster* = 2 adalah 0.729723 dengan rata-rata 0.672942. Di dalam penelitian ini terdapat saran yaitu perlunya riset lanjutan untuk menentukan nilai k dengan lebih banyak proses percobaan pengukuran tingkat akurasi dan kemurnian *cluster* (Husni dkk, 2015).

Penelitian yang terkait lainnya yaitu penelitian yang dilakukan oleh Alfina dkk ditahun 2012 yang berkaitan dengan problem kerja praktek jurusan teknik industri ITS pada tahun 2011. Semua data yang sudah diperoleh, dipilih *keyword-keyword* yang dapat mewakili problem kerja praktek jurusan teknik industri ITS pada tahun 2011. *Keyword* yang dipilih akan digunakan untuk membentuk matriks metadata yang menunjukkan frekuensi dari setiap *keywords* dalam setiap problem yang disampaikan oleh mahasiswa yang mengambil kerja praktek pada tahun 2011. Diperoleh kesimpulan:

- a. Kombinasi algoritma *hierarchical clustering* dan *kmeans* menghasilkan pengelompokan data yang lebih baik jika dibandingkan dengan *k-means* dalam semua pengujian.
- b. Dengan evaluasi koefisien *cophenetic*, metode *clustering* terbaik dihasilkan oleh *average linkage clustering*.
- c. Dalam studi kasus Problem Kerja Praktek Jurusan Teknik Industri ITS, dari kombinasi *hierarchical clustering* dan *K-means* yang ada, kombinasi *single linkage clustering* dan *K-means* menghasilkan pengelompokan data yang terbaik dibandingkan dengan metode hierarki yang lainnya. (Alfina dkk, 2012)

Landasan teori yang digunakan dalam penelitian ini adalah penjelasan teori dari buku yang berhubungan dengan penelitian ini, diantaranya:

1. Pengertian Data Mining

Data mining disebut juga *knowledge discovery* karena merupakan bidang yang berupaya untuk menemukan informasi yang punya arti dan berguna dari jumlah data yang besar. Data mining merupakan suatu proses yang interaktif atau terotomatisasi untuk menemukan pola (*pattern*) data tersebut dan memprediksi kelakuan (*trend*) di masa mendatang berdasarkan pola data tersebut. Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu (Nofriansyah, 2014).

2. Pengertian Analisis *Cluster*

Analisis *cluster* adalah teknik yang digunakan untuk mengklasifikasikan objek-objek atau kasus-kasus menjadi kelompok (*cluster*) yang relatif homogen. Dengan menggunakan analisis *cluster*, sejumlah data yang berbeda akan diklasifikasikan ke dalam satu atau lebih *cluster*. Setiap *cluster* kemudian akan berisi objek yang memiliki kemiripan atau karakteristik yang sama, dan objek yang berlainan *cluster* tidak mirip satu sama lain. Pengelompokan data bisa menggunakan metode *Non-Hierarchical Cluster* atau *Hierarchical Cluster*. Pada *Non-Hierarchical Cluster*, peneliti menentukan terlebih dahulu jumlah kelompok atau *cluster* yang diinginkan. Setelah itu, proses *clustering* dilakukan untuk mengetahui karakteristik (kemiripan) yang ada pada masing-masing *cluster* (Hidayat dan Istiadah, 2011).

Landasan teori lainnya yang digunakan dalam penelitian ini adalah penjelasan teori yang diambil dari materi kuliah dan jurnal yang berhubungan dengan penelitian ini, diantaranya:

1. Pengertian *Agglomerative Hierarchical Clustering* (AHC)

Metode ini dimulai dengan setiap objek dinyatakan sebagai kluster tersendiri. Kedekatan pada kluster baru dihitung ulang dan kluster paling dekat digabung lagi. Proses tersebut dilakukan secara berulang sampai seluruh data (objek) menjadi 1 kluster (Putra, 2010).

Langkah-langkah AHC (Supianto, 2014) antara lain:

- a. Hitung matrik jarak antar data.
- b. Ulangi langkah 3 dan 4 hingga hanya satu kelompok yang tersisa.
- c. Gabungkan dua kelompok terdekat berdasarkan parameter kedekatan yang ditentukan.
- d. Perbaharui matrik jarak antar data untuk mempresentasikan kedekatan diantara kelompok baru dan kelompok yang masih tersisa.
- e. Selesai.

2. Pengertian *K-means Clustering*

Menurut Sarwono *algoritma-k-means* adalah metode yang mempartisi data ke dalam cluster (kelompok) sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam cluster yang sama dan data yang mempunyai karakteristik berbeda dikelompokkan ke dalam kelompok yang lain (Sarwono).

Langkah-langkah k-means clustering (Sarwono) antara lain:

- a. Menentukan k sebagai jumlah kluster yang ingin dibentuk.
 - b. Membangkitkan nilai random untuk pusat *cluster* awal (*centroid*) sebanyak k.
- Menghitung jarak setiap data input terhadap masing-masing *centroid* menggunakan rumus jarak *Euclidean* (*Euclidean Distance*) hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*. Berikut adalah persamaan

c. *Euclidean Distance* :

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2} \quad (1)$$

Dimana:

d = jarak

i = indeks dari *cluster*

j = indeks dari variabel

x_i = data kriteria ke- i

μ_j = *centroid* pada *cluster* ke- j

d. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil).

e. Memperbaharui nilai *centroid*. Nilai *centroid* baru diperoleh dari rata-rata *cluster* yang bersangkutan dengan menggunakan rumus :

$$\mu_j(t+1) = \frac{1}{N_{S_j}} \sum_{j \in S_j} x_j \quad (2)$$

Dimana :

j = indeks dari variabel

$\mu_j(t+1)$ = *centroid* baru pada iterasi ke- j

N_{S_j} = banyak data pada *cluster* S_j

S_j = *Silhouette* data ke- j

x_j = data kriteria ke- j

f. Melakukan perulangan dari langkah 2 hingga 5, sampai anggota tiap *cluster* tidak ada yang berubah.

Jika langkah 6 telah terpenuhi, maka nilai pusat *cluster* (μ_j) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.

METODOLOGI PENELITIAN

Dalam penelitian ini digunakan data berita berupa teks berbahasa Indonesia, dimana peneliti mengambil beberapa sampel data berita melalui situs www.suara.com. Dengan format data yaitu koleksi dokumen berita 1 terdapat data berita berjumlah 8 data berita, koleksi dokumen berita 2 terdapat data berita berjumlah 10 data berita, dan koleksi dokumen berita 3 terdapat data berita berjumlah 12 data berita, masing-masing data berita berupa teks. Untuk metode pengumpulan data dalam penelitian ini yaitu:

1. Studi Pustaka

Mencari, membaca, dan mempelajari buku, jurnal dan situs-situs yang berkaitan dari internet serta referensi lain yang terkait.

2. Dokumen

Pengambilan data melalui dokumen tertulis maupun elektronik dari lembaga atau institusi. Dokumen diperlukan untuk mendukung kelengkapan data yang lain.

Alat yang digunakan dalam penelitian ini adalah perangkat keras (*hardware*) dan perangkat lunak (*software*) dengan spesifikasi sebagai berikut:

1. *Hardware* : Processor : Intel(R) i3 2.00 Ghz, RAM : 4 GB, Harddisk : 500GB, dan Layar : LCD 14".
2. *Software* : Windows 10 Pro, Java dengan NetbeansIDE versi 8.2, XAMPP versi 3.2.2, dan Software Ideas Modeler versi 10.68.

Pada tabel 2 contoh data berita merupakan sampel data berita dari masing-masing *cluster*, terdiri dari 4 *cluster* yaitu cluster 0 (kategori berita bisnis), cluster 1 (kategori berita bola), cluster 2 (kategori berita teknologi), dan cluster 3 (kategori berita otomotif). Diambil sampel data berjumlah 4 data berita untuk masing-masing *cluster*. Masing-masing data akan diberikan inisial *cluster*. Inisialisasi *cluster* dapat dilihat pada tabel 1.

Tabel 1 Inisialisasi *cluster*

Kategori	Cluster
Berita Bisnis	0
Berita Bola	1
Berita Teknologi	2
Berita Otomotif	3

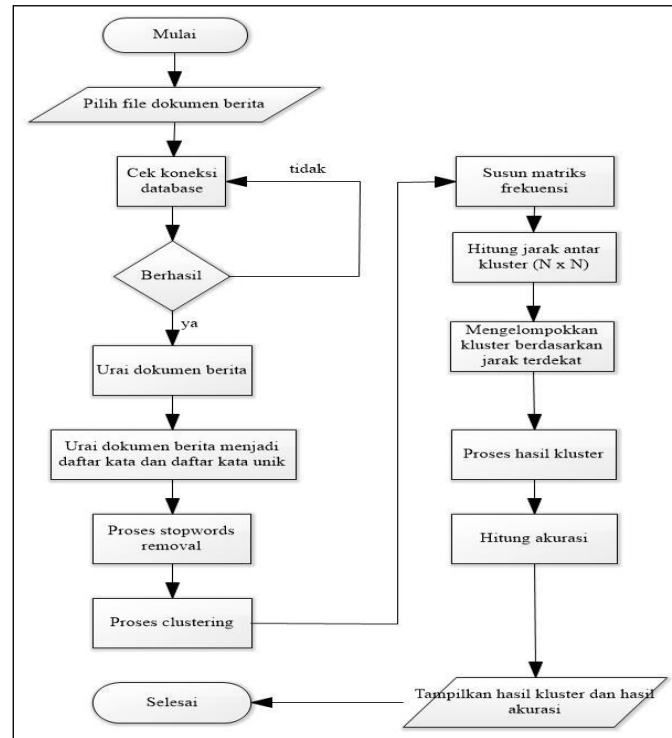
Tabel 2 Contoh data berita

No	Isi	Label
1	<p>Pemerintah dan para pelaku usaha dari negara-negara Samudera Hindia akan bertemu pada Indian Ocean Rim Association (IORA) Summit. IORA Summit akan berlangsung pada 5-7 Maret 2017 di Jakarta Convention Center, Jakarta.</p> <p>Menteri Perdagangan Enggartiasto Lukita menjelaskan IORA memiliki peran yang sangat strategis sebagai forum pendorong stabilitas kawasan dan IORA merupakan masa depan ekonomi di dunia.</p> <p>"Saat ini menjadi momentum yang tepat mengingat pertumbuhan perekonomian di beberapa negara anggota IORA terbilang tinggi. Peningkatan aktivitas perdagangan dan investasi IORA juga dapat semakin mendorong pertumbuhan ekonomi yang adil dan berkelanjutan, serta menciptakan lapangan kerja baru," ungkap Mendag Enggar dalam siaran pers Kemendag, Sabtu (4/3/2017).</p> <p>Perdagangan intra-regional IORA di tahun 2015 mencapai 777 miliar dolar AS atau naik 300 persen dibandingkan tahun 1994 yang sebesar 233 miliar dolar AS. Selain itu, Samudera Hindia merupakan 70 persen jalur perdagangan dunia, termasuk jalur distribusi minyak dan gas. Bahkan lebih dari setengah kapal kontainer dan dua per tiga kapal tanker minyak dari seluruh dunia melewati kawasan ini.</p> <p>IORA mencakup kurang lebih 2,7 miliar penduduk (35 persen dunia). Namun, perannya baru 12 persen pangsa pasar dunia, 10 persen PDB global, dan 13 persen tujuan penanaman modal asing (PMA). Sebesar 96 persen perdagangan intra-IORA dikuasai enam negara, yaitu Singapura, Malaysia, India, Indonesia, Australia, dan Afrika Selatan. Beberapa negara yang tengah menjadi perhatian penting Pemerintah Indonesia dalam hal perdagangan adalah Bangladesh, Kenya, Mozambik, Afrika Selatan, Uni Emirat Arab, dan Iran.</p> <p>"Bagi Kemendag, IORA sangat strategis dan sejalan dengan strategi diversifikasi pasar tujuan ekspor, hal ini sesuai dengan arahan Bapak Presiden pada pembukaan Rakernas Perdagangan bulan lalu. Kemendag akan mengoptimalkan pertemuan ini untuk melakukan ekspansi atau pendalaman terhadap pasar-pasar baru yang potensial," tegas Mendag.</p> <p>Saat ini Indonesia tengah menjalin kesepakatan perdagangan bebas dengan lima negara anggota IORA (Malaysia, Singapura, Thailand, India dan Australia) dan satu negara mitra wicara IORA (Jepang) dalam kerangka ASEAN dan ASEAN Plus.</p>	0

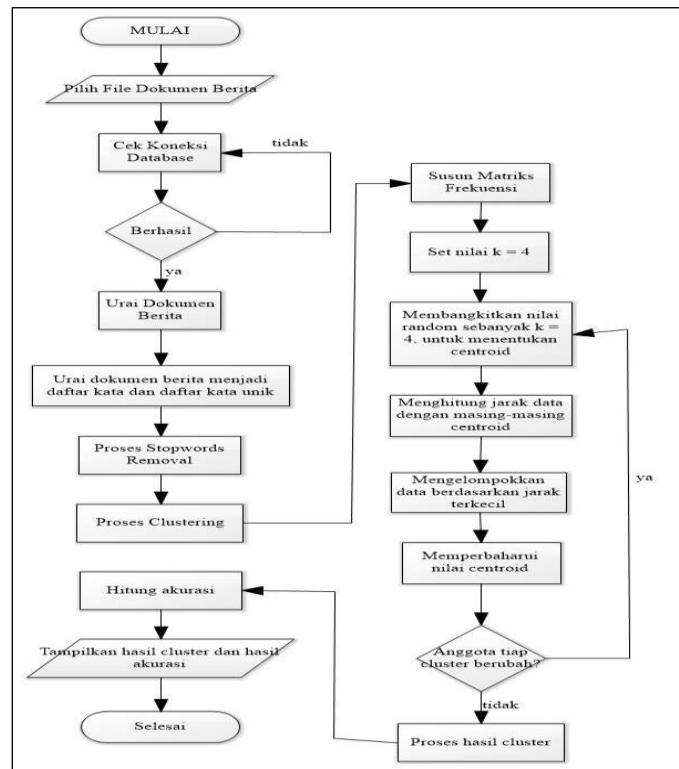
	<p>"Melalui IORA, Indonesia dapat menjajaki peningkatan cakupan jalinan kerja sama perdagangan bilateral dan multilateral dengan beberapa negara kunci, seperti Uni Emirat Arab, Afrika Selatan, Bangladesh dan Iran, serta Amerika Serikat dan Inggris," tutur Mendag.</p> <p>IORA merupakan forum kerja sama antarnegara terbesar di Samudera Hindia yang berdiri pada tahun 1997. IORA beranggotakan 21 negara yaitu Australia, Afrika Selatan, Bangladesh, Komoros, India, Indonesia, Iran, Kenya, Madagaskar, Malaysia, Mauritius, Mozambik, Oman, Seychelles, Singapura, Somalia, Sri Lanka, Tanzania, Thailand, Uni Emirat Arab, dan Yaman; serta 7 negara mitra wicara yaitu Amerika Serikat, Inggris, Jepang, Jerman, Mesir, Tiongkok, dan Prancis.</p> <p>Empat negara anggota IORA (Afrika Selatan, Australia, India, dan Indonesia) serta 6 negara mitra IORA (AS, RRT, Jerman, Inggris, Jepang dan Perancis) merupakan anggota G20.</p> <p>"Hal ini membuktikan betapa strategisnya peran IORA dalam perekonomian dunia," imbuh Mendag.</p> <p>Penyelenggaraan IORA Summit 2017 merupakan salah satu gagasan dan prakarsa strategis Indonesia pada masa keketuaan Indonesia pada IORA periode 2015-2017. Selain itu, Indonesia juga mengagagas pembentukan IORA Concord sebagai outcome strategis 20 tahun IORA.</p> <p>Indonesia secara resmi menjadi Ketua IORA untuk periode 2015–2017 pada the 15th IORA Council of Ministers Meeting di Padang. Tema keketuaan Indonesia di IORA adalah Strengthening Maritime Cooperation for a Peaceful, Stable and Prosperous Indian Ocean.</p>	
2	<p>Ahmad Nur Hardianto menjadi bintang dalam game internal seleksi tahap II tim nasional U-22 yang berlangsung di lapangan Universitas Pelita Harapan, Karawaci, Tangerang, Kamis (2/3/2017). Pemain Persela Lamongan itu mencetak dua gol di laga tersebut.</p> <p>Usai pertandingan, Ahmad yang membela kubu rompi kuning tidak ingin jumawa. Pemain 21 tahun itu mengatakan hanya memanfaatkan peluang yang ada.</p> <p>"Yang penting saya buktikan lagi saja. Manfaatkan peluang yang ada di depan kotak penalti," kata Dian di Hotel Yasmin, Karawaci, Tangerang.</p> <p>Dalam pertandingan yang berlangsung selama 2x30 menit, gol tercipta di babak pertama. Sementara di paruh kedua tidak ada tambahan gol.</p> <p>Berkat penampilan gemilangnya, bukan tidak mungkin sang pemain lolos seleksi timnas U-22 Indonesia. Terlebih lagi, persaingan di lini depan tidak begitu padat.</p>	1
3	<p>Motorola positif akan merilis Moto M dalam waktu dekat. Hal ini diungkap Anvid Erdian, 4P Manager MBG Lenovo Indonesia, perusahaan yang telah mengakuisi Motorola.</p> <p>"Kami akan membawa Moto M ke pasar Indonesia sebentar lagi. Tapi belum bisa menginformasikan kapan pastinya," katanya dalam ajang Motorola Indonesia Media Camp Modsventure in a Snap! di Bandung, Rabu malam (15/3/2017).</p> <p>Kabarnya, Moto M akan mengisi pasar yang belum pernah dimiliki Motorola Indonesia. Meskipun Anvid tidak memaparkan tanggal peluncuran resmi dan harga yang akan dibanderol di pasar Indonesia, tapi dia membocorkan beberapa spesifikasi ponsel anyar ini.</p>	2

	<p>"Konsumen Indonesia menurut studi dalam memilih hp berdasarkan desain. Top key decision untuk purchase device," ujar dia.</p> <p>Secara desain, Anvid memaparkan, bodi Moto M akan dibalut bahan metal dengan kualitas tinggi, unibody, 5,5 inci full HD, RAM 4GB dan Quadcore Pocessor memperkuat kinerja Moto M.</p> <p>Ponsel ini sudah mencakup jaringan 4G plus dengan kecepatan hingga 300 Mbps. Soal kamera ponsel anyar itu dilengkapi sensor 8MP untuk kamera depan dan 16MP di kamera belakang.</p> <p>Moto M mengoperasikan Android 6.0 Mashmallow dengan penyimpanan internal 32GB yang dapat diperluas hingga 128GB. Handset tersebut telah disematkan baterai 3050 mAh.</p> <p>Soal teknologi, Moto M menambah fitur pemindai sidik jari dan mampu bertahan dengan percikan air.</p>	
4	<p>Volkswagen (VW) memamerkan sebuah mobil konsep yang tampaknya akan menjadi suksesor minibus legendaris mereka, Combi, di masa depan. Kendaraan purwarupa tersebut bernama Sedric Concept dan merupakan konsep swakemudi pertama VW.</p> <p>Sedric Concept, yang merupakan kependekan dari 'Self-driving Car', dipamerkan untuk pertama kali Sedric Concept di depan publik pada malam inagurasi jelang Geneva Motor Show 2017 yang berlangsung mulai hari ini, demikian dilaporkan laman daring Motor1 pada Selasa (7/3/2017).</p> <p>Sedric Concept dikembangkan oleh VW Group Future Center bersama-sama dengan VW Group Research di Jerman. Minibus futuristik ini memiliki level tertinggi dari teknologi swakemudi.</p> <p>Karena itulah, Sedric Concept tidak memiliki setir, pedal gas dan rem, serta kokpit. Kabinnya benar-benar disediakan bagi penumpang yang tinggal duduk, memberi perintah, dan menikmati perjalanan.</p> <p>Sedric Concept dirancang sebagai moda transportasi pribadi yang 100 persen otomatis, terkoneksi, dan nihil emisi karena disokong baterai listrik.</p> <p>Dengan hanya menekan sebuah tombol, Sedric Concept akan menjemput pemiliknya. Ia juga dirancang sebagai asisten pribadi digital yang dapat berdiskusi mengenai jalur untuk sampai ke tujuan, estimasi waktu sampai, hingga kondisi lalu lintas.</p> <p>Segala informasi perjalanan akan ditampilkan lewat layar OLED besar berteknologi augmented reality di kaca depan. Layar ini juga menjadi bagian dari sistem infotainment bagi penumpang.</p> <p>Adapun konfigurasi kursi Sedric Concept ialah 2+2. VW sendiri tak menjelaskan rencana realisasi mobil ini. Namun, melihat sebegitu canggih teknologi di dalamnya, kemungkinan besar VW butuh waktu cukup lama jikapun benar-benar ingin mewujudkannya.</p> <p>Bersamaan dengan dipamerkannya Sedric Concept, VW juga mengungkapkan lagi rencana untuk mengaspalkan tak kurang dari 30 mobil listrik hingga 2025.</p>	3

Diagram alir (*flowchart*) dapat dilihat pada gambar 1 untuk sistem program AHC dan gambar 2 untuk sistem program *k-means clustering*.



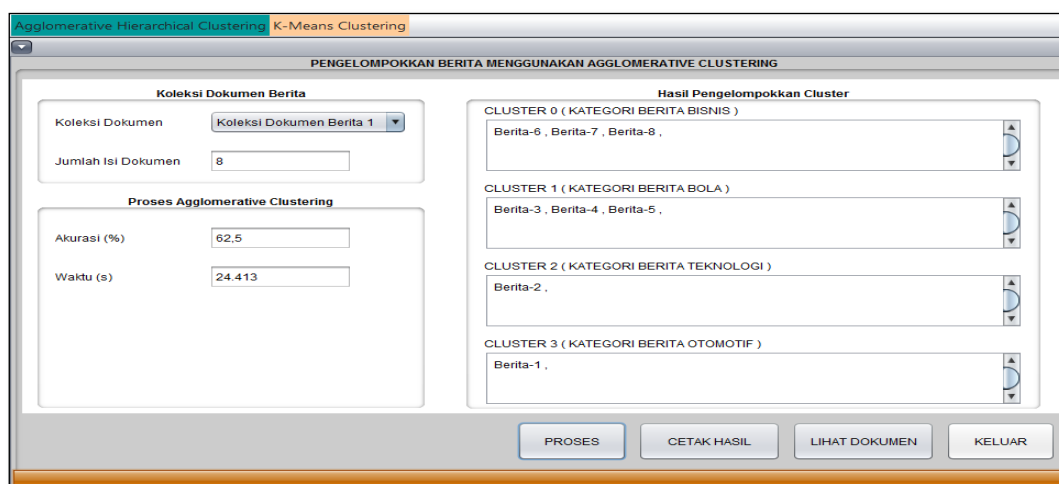
Gambar 1 *Flowchart* sistem program AHC



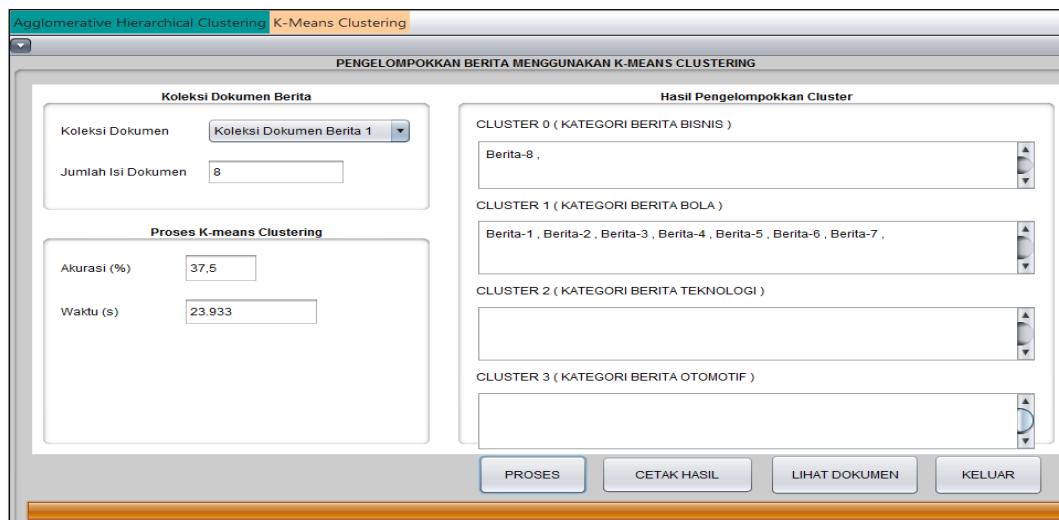
Gambar 2 Flowchart sistem program k-means clustering

PEMBAHASAN

Untuk melakukan pengelompokkan data berita terlebih dahulu memilih pilihan koleksi dokumen berita, kemudian diklik button proses untuk melakukan proses pengelompokkan data berita. Berikut ini adalah tampilan hasil pengelompokkan berita menggunakan AHC koleksi dokumen berita 1 yang dapat dilihat pada gambar 3 dan tampilan hasil pengelompokkan berita menggunakan k-means clustering koleksi dokumen berita 1 yang dapat dilihat pada gambar 4.



Gambar 3 Hasil pengelompokkan berita dengan AHC koleksi dokumen berita 1



Gambar 4 Hasil pengelompokkan berita dengan *k-means clustering* koleksi dokumen berita 1

Untuk mengetahui kinerja metode-metode tersebut dalam mengelompokkan data berita, maka pengelompokkan dilakukan dengan cara menghitung nilai akurasi dari pengelompokkan berita. Akurasi adalah ketepatan hasil dari pengelompokkan data, dengan perhitungan jumlah berita benar yang sudah dikluster dibagi dengan jumlah total data berita dikali 100%. Rumus perhitungannya sebagai berikut:

$$\text{Akurasi} = \frac{\text{hasil data pengelompokkan}}{\text{jumlah total data berita}} \times 100\% \tag{3}$$

Untuk hasil akurasi pengelompokkan berita selengkapnya dapat dilihat pada tabel 3. Akurasi yang diperoleh akan semakin menurun, apabila data berita pada koleksi dokumen berita semakin banyak jumlahnya.

Tabel 3 Hasil Akurasi Klasifikasi Berita

No	Metode Clustering	Akurasi (%)			Waktu (s)		
		Koleksi Dokumen Berita 1	Koleksi Dokumen Berita 2	Koleksi Dokumen Berita3	Koleksi Dokumen Berita 1	Koleksi Dokumen Berita 2	Koleksi Dokumen Berita3
1	Agglomerative Hierarchical Clustering	62.5%	50%	33.33%	24.413 s	30.679 s	33.296 s
2	<i>K-means Clustering</i>	0%	10%	0%	24.259 s	30.687 s	35.761 s
		12.5%	20%	8.33%	24.256 s	30.695 s	34.736 s
		25%	30%	16.67%	24.221 s	30.648 s	34.792 s
		37.5%	40%	25%	23.933 s	31.712 s	35.999 s
		50%	50%	25%	26.754 s	30.687 s	33.419 s
		62.5%	60%	33.33%	24.239 s	30.652 s	36.128 s
Jumlah rata-rata		31.25%	35%	18.06%	24.61 s	30.85 s	35.139 s

Total rata-rata akurasi (%) <i>agglomerative clustering</i>	49%	Total rata-rata waktu (s) <i>agglomerative hierarchical clustering</i>	29.5 s
Total rata-rata akurasi (%) <i>k-means clustering</i>	28.103%	Total rata-rata waktu (s) <i>k-means clustering</i>	30,19 s

Dari tabel 3 akurasi klasifikasi berita terdapat 3 koleksi dokumen berita yaitu: koleksi dokumen berita 1, koleksi dokumen berita 2, dan koleksi dokumen berita 3. Menggunakan 2 metode *clustering*, metode yang pertama adalah AHC, hanya diambil 1 sampel hasil akurasi, karena AHC prosesnya tidak secara random (acak) sehingga hasil akurasi tidak akan berubah-ubah. Metode yang kedua menggunakan metode *k-means clustering*, dimana masing-masing koleksi dokumen berita diambil 6 sampel hasil akurasi dari pengelompokan berita, karena *k-means clustering* menggunakan proses secara random (acak) sehingga hasilnya akan berubah-ubah.

KESIMPULAN

1. Dalam penelitian ini sudah dapat mengelompokkan data-data berita pada dokumen berita secara otomatis.
2. Jika hasil akurasi dilihat dari uji coba satu persatu, maka hasil akurasi metode *k-means clustering* lebih baik sedikit dari metode *agglomerative hierarchical clustering*. Ini dapat dibuktikan dengan melihat hasil akurasi menggunakan *k-means clustering* pada koleksi dokumen berita 2 diperoleh hasil akurasi yaitu 60%, sedangkan AHC diperoleh hasil akurasi yaitu 50.
3. Jika hasil penelitian dilihat dari total rata-rata eksekusi waktu yang diperoleh masing-masing metode juga tidak terdapat perbedaan yang signifikan. Menggunakan metode *k-means clustering* dihasilkan total rata-rata waktu yaitu 30,14 s, sedangkan menggunakan metode AHC dihasilkan total rata-rata waktu yaitu 30,2 s. Sehingga jika diambil total rata-rata waktu, maka *k-means clustering* akan lebih cepat sedikit dari pada menggunakan AHC.
4. Jadi dari hasil penelitian metode yang lebih bagus digunakan adalah metode AHC, karena akurasi dan waktu untuk mengeksekusi program tidak perlu diulang-ulang, sehingga lebih menghemat waktu pengelompokan data berita. Sedangkan metode *k-means clustering* membutuhkan waktu lebih banyak, karena diproses secara random dan harus mengulang-ulang untuk mendapatkan akurasi yang paling tinggi.
5. Akurasi yang diperoleh akan semakin menurun, apabila data berita pada koleksi dokumen berita semakin banyak jumlahnya.
6. Dari hasil penelitian ini, dapat disimpulkan bahwa pengelompokan data berita dengan hasil akurasi 60 % ke atas sudah lumayan bagus akurasi/ketepatan dalam pengelompokan data berita, sedangkan untuk hasil akurasi 60 % ke bawah akurasi/ketepatan dalam pengelompokan data berita belum terlalu bagus.

DAFTAR PUSTAKA

- Romli, A. S. M. 2012. *Jurnalistik online: Panduan Mengelola Media Online*. Nuansa. Bandung.
- Hidayat, T., dan Istiadah, N. 2011. *Panduan Lengkap SPSS 19 untuk Mengolah Data Statistik Penelitian*. Media Kita. Jakarta.
- Atmajaya, D. 2016. *K-means Algorithm*. Fakultas Ilmu Komputer UMI. Makassar.
- Kusumawardani, Y. 2016. *Klasifikasi Opini Menggunakan Metode K-mean Clustering*. Kerja Praktek. IST Akprind Yogyakarta
- Husni., Negara, Y. D. P., dan Muhammad, S. 2015. *Clusterisasi Dokumen Web (Berita) Bahasa Indonesia Menggunakan Algoritma K-means*. Jurnal Simantec vol. 4, no. 3. Madura.

- Alfina, T., Santosa, B., dan Barakbah, A, R. 2012. *Analisa Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data (Studi Kasus: Problem Kerja Praktek Jurusan Teknik Industri ITS)*. Jurnal Teknik ITS Vol. 1. Surabaya.
- Nofriansyah, D. 2014. *Konsep Data Mining Sistem Pendukung Keputusan*. Deepublish. Yogyakarta.
- Putra, D. 2010. *Pengolahan Citra Digital*. Andi Offset. Yogyakarta.
- Supianto, A. 2014. *Pengenalan Pola Hierarchical Clustering*. Materi PPT PTIIK. Universitas Brawijaya.
- Sarwono, Y. T. *Aplikasi Model Jaringan Syaraf Tiruan Dengan Radial Basis Function Untuk Mendeteksi Kelainan Otak (Stroke Infark)*. Jurnal Sistem Informasi