

Analisis Perbandingan Algoritme *K-Means* dan *Isodata* untuk Klasterisasi Data Kejadian Titik Api di Wilayah Sumatera pada Tahun 2001 hingga 2014

Edo Fadila Sirat¹, Budi Darma Setiawan², Fatwa Ramdani³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹edofadila@gmail.com, ²s.budidarma@ub.ac.id, ³fatwaramdani@ub.ac.id

Abstrak

Fenomena kebakaran merupakan fenomena yang tidak asing di Indonesia. Tingginya angka kejadian kebakaran yang terjadi di Indonesia membutuhkan perhatian khusus dari pemerintah, agar setiap bencana alam seperti kebakaran hutan dapat ditanggulangi. Hasil pantauan satelit tercatat pada sebuah file data titik api dengan ukuran data yang cukup besar sehingga data sulit diolah untuk menjadi informasi yang mudah diterima oleh pengguna. Berdasarkan data yang diperoleh dari situs EOSDIS tercatat sebanyak 289.256 kejadian titik api di wilayah Sumatera dalam rentan waktu antara 2001 hingga 2014. Dibutuhkan sebuah algoritme untuk melakukan segmentasi data atau klasterisasi data, agar data yang besar dapat diolah menjadi sebuah informasi yang baik bagi pengguna. Dalam penelitian ini dilakukan studi perbandingan algoritme klasterisasi antara *K-Means* dan *Isodata*. Kedua algoritme yang digunakan dalam penelitian ini dinilai berdasarkan kualitas kluster yang dihasilkan. Metode yang digunakan dalam mengukur kualitas kluster dalam penelitian ini adalah *Silhouette Coefficient* (SC). Hasil akhir nilai SC algoritme *K-Means* sebesar 0.999997187 dan algoritme *Isodata* sebesar 0.999957161, sehingga dalam hal ini, algoritme *K-Means* memiliki nilai SC yang lebih tinggi dibandingkan dengan algoritme *Isodata* dalam mengkluster data kejadian titik api dengan selisih nilai SC yang kecil.

Kata kunci: *k-means, isodata, klasterisasi, titik api, titik panas, analisis*

Abstract

*Fire phenomenon is a familiar phenomenon in Indonesia. The high number of fire incidents in Indonesia requires special attention from the government, so that any natural disasters such as forest fires can be overcome. Satellite monitoring results are recorded on a data file of fire points with a large enough data numbers so that the data is difficult to be processed to become information that is easily received by the user. Based on data obtained from the EOSDIS site recorded as many as 289,256 fire spots occurrence in the region of Sumatra in the timeframe between 2001 to 2014. It takes an algorithm to segment the data or cluster the data, so that large data can be processed into a good information for the user. In this study a comparative study of clustering algorithms between *K-Means* and *Isodata* was conducted. Both algorithms used in this study were assessed based on the quality of the clusters produced. The algorithm used in measuring the quality of cluster in this research is *Silhouette Coefficient* (SC). The final result value of *Silhouette Coefficient* *K-Means* method is 0.999997187 and *Isodata* method is 0.999957161, so in this case, *K-Means* algorithm has a higher SC value compared to the *Isodata* algorithm in clustering the data of fire spots with a small SC value difference.*

Keywords: *k-means, isodata, clustering, fire point, hotspot, analysis*

1. PENDAHULUAN

Fenomena kebakaran merupakan fenomena yang tidak asing di Indonesia. Maraknya kejadian seperti kebakaran hutan menambah daftar panjang peristiwa kejadian kebakaran di

beberapa wilayah Indonesia. Terdapat beberapa faktor yang menyebabkan terjadinya peristiwa kebakaran hutan. Peristiwa kebakaran dapat terjadi dikarenakan faktor alami maupun faktor perbuatan manusia (Rasyid, 2014).

Pada umumnya faktor alami yang menjadi pemicu kejadian peristiwa kebakaran adalah cuaca terik pada musim kemarau hingga terjadi kekeringan yang berdampak pada lahan gambut (Harrison, et al., 2009). Secara umum Kerkeringan yang terjadi di Indonesia dikarenakan Iklim di Indonesia sangat dipengaruhi oleh El Nino. Kebakaran yang terjadi Indonesia, banyak membakar lahan gambut yang mengering. Lahan gambut yang terbakar seringkali terjadi disebabkan oleh perbuatan manusia (Harrison, et al., 2009).

Tingginya angka kejadian kebakaran yang terjadi di Indonesia membutuhkan perhatian khusus dari pemerintah, agar setiap bencana alam seperti kebakaran hutan maupun bencana alam lainnya dapat ditanggulangi. Terjadinya peristiwa kebakaran dapat dipantau melalui satelit. Hasil pantauan satelit tercatat pada sebuah file data titik api dengan jumlah data yang cukup besar sehingga data sulit diolah dan dipahami untuk menjadi informasi yang mudah terima oleh pengguna. Dibutuhkan sebuah metode untuk melakukan segmentasi data agar data dengan jumlah yang besar dapat diolah menjadi sebuah informasi yang baik bagi pengguna. Ada beberapa metode yang dapat digunakan untuk melakukan segmentasi data.

Pada penelitian sebelumnya yang dilakukan oleh Muningsih dan Kiswati (2015) dengan membahas tentang pengelompokan produk yang dijual pada *online shop* ragam Jogja menjadi beberapa kluster untuk mengetahui produk mana yang paling banyak diminati sehingga dapat dilakukan penyesuaian dengan stok. Variabel yang digunakan adalah kode produk, jumlah transaksi, volume penjualan dan rata rata penjualan. Pengukuran dan evaluasi dalam penelitian ini menghasilkan 3 kluster dengan tingkat akurasi tinggi pada kluster 1 mencapai 80%, kluster 2 mencapai 60%, dan kluster 3 mencapai 80 % (Muningsih & Kiswati, 2015).

Pada penelitian yang dilakukan oleh Ghosh dan Dubey (2013), membahas tentang analisis perbandingan dua algoritme yaitu *K-Means* dan Fuzzy *C-Means* dalam melakukan analisis data. Penelitian ini diuji berdasarkan kompleksitas waktu komputasi antara dua algoritma. Algoritme *K-Means* menghasilkan waktu 0.443755 detik sedangkan pada algoritme FCM menghasilkan waktu 0.781679 detik (Ghosh & Dubey, 2013).

Sementara itu, pada penelitian yang dilakukan oleh Yu dkk (2012), membahas tentang algoritme *Fuzzy Isodata* untuk penilaian

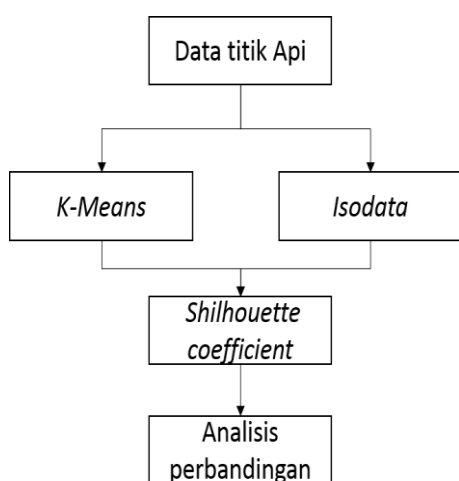
resiko kredit dari *Emerging Information Technology (EIT)*. Dalam penelitian ini terdapat beberapa indikator sebagai parameter penilaian. Indikator penilaian dalam penelitian ini terdiri dari 4 indikator utama dan 15 indikator sekunder. Empat indikator utama yaitu ilmiah, operasional, prinsip objektif serta kombinasi kuantitatif dan indeks evaluasi. Metode yang digunakan dalam penelitian ini adalah *Fuzzy Isodata*. Hasil yang diperoleh dari klasterisasi algortima *Isodata* menghasilkan 10 investigasi perusahaan EIT harus ditetapkan menjadi 5 kredit nilai sehingga dalam penelitian ini algoritme *Isodata* merupakan hal yang tepat apabila diterapkan pada perusahaan EIT (Yu, et al., 2012).

Pada penelitian yang dilakukan oleh Liu dkk (2012), membahas tentang pemilihan fitur dari data ekspresi gen berdimensi tinggi. Dalam penelitian ini data yang digunakan adalah data dalam bidang bioinformatika. Dalam penelitian ini, metode yang diusulkan adalah algoritma *Fuzzy Isodata*. Evaluasi hasil dalam penelitian ini, dilakukan dengan 5 *microarray dataset* diantaranya, *DLBCL*, *Acute Leukemia*, *Multiple myeloma*, *Colon* dan *Prostate*. Hasil yang diperoleh menunjukkan bahwa fitur yang dipilih memiliki kinerja yang baik dalam melakukan klasifikasi dan klasterisasi pada 5 gen dataset profil ekspresi, sehingga dalam penelitian ini disimpulkan bahwa percobaan yang dilakukan dengan *isodata RFE* mampu dalam memilih fitur dengan klasifikasi yang kuat dan kemampuan klasterisasi (Liu, et al., 2012).

Berdasarkan beberapa penelitian sebelumnya, metode *K-Means* dan *Isodata* dapat diterapkan pada berbagai macam studi kasus dengan cukup baik. Pada metode *K-Means* mendapatkan penilaian yang cukup apabila ditinjau dari segi kualitas kluster dan waktu komputasi, sedangkan metode *isodata* mendapatkan nilai yang cukup baik apabila ditinjau dari segi kualitas kluster. Dalam penelitian ini kedua metode tersebut akan diterapkan untuk klasterisasi data kejadian titik api wilayah Sumatera pada tahun 2001 hingga 2014. Pada penelitian ini fokus masalah yang dibahas adalah bagaimana kualitas kluster metode *K-Means* dan *Isodata* dan analisis hasil SC antara keduanya pada data kejadian titik api.

2. METODOLOGI DAN PERANCANGAN

Tahapan yang dilakukan dimulai dari input data kejadian titik api pada masing masing sistem. Sistem yang digunakan adalah sistem klasterisasi dengan metode *K-Means* dan *Isodata*. Hasil klaster yang didapat dari masing masing sistem diolah dengan metode *Shilhouette Coefficient* (SC) untuk mendapatkan nilai kualitas klaster masing masing metode. Analisis perbandingan dilakukan berdasarkan nilai SC yang didapat dari masing masing metode. Gambar 1 merupakan diagram metodologi pada penelitian ini .



Gambar 1. Metodologi

2.1. Pengumpulan Data

Data yang digunakan adalah data kejadian titik api pada wilayah Sumatera. Data didapat dari situs EOSDIS dengan alamat web earthdata.nasa.gov. Data Kejadian titik api merupakan data yang bersifat publik dan dapat diperoleh secara gratis. Gambar 2 merupakan peta wilayah yang diseleksi untuk studi kasus data kejadian titik api pada wilayah Sumatera pada tahun 2001 hingga 2014.



Gambar 2. Peta seleksi wilayah.

Gambar 2 memvisualisasikan peta wilayah yang di seleksi, yaitu pulau Sumatera. Data

kejadian titik api yang di pilih terhitung dari tanggal 1 Januari 2001 hingga 31 Desember 2014. Data kejadian titik api yang diperoleh terdiri dari 289,257 baris data dan 13 kolom. Variable 13 kolom yang didapat adalah *geom*, *latitude*, *longitude*, *brigness*, *scan*, *track*, *acq_date*, *acq_time*, *satellite*, *confidence*, *version*, *bright_t31*, *frp*.

2.2. Perancangan Sistem

Perancangan sistem merupakan tahapan untuk melakukan pemodelan berdasarkan studi kasus yang diangkat dalam penelitian ini. Pemodelan dilakukan dengan pembuatan diagram alir, manualisasi algortime *K-Means* dan manualisasi algoritme *Isodata*. Berdasarkan 13 variable yang didapat dipilih lima variabel yang akan di klaster yaitu *latitude*, *longitude*, *brightness*, *acq_date*, *convidence*.

Latitude dalam bahasa Indonesia adalah garis bujur. Garis bujur adalah garis khayal yang ditarik dari kutub utara ke kutub selatan. Garis bujur dapat membagi bumi menjadi dua bagian, diantaranya bumi bagian timur dan bumi bagian barat. Garis bujur dapat menentukan sebuah lokasi tertentu. Hal ini, dikombinasikan dengan adanya garis lintang. Dalam koordinat garis lintang dingambarkan dengan sumbu x, sedangkan garis bujur dilambangkan sebagai sumbu y. Garis bujur juga dapat dijadikan sebagai dasar dalam menentukan pembagian waktu di seluruh bagian bumi (Sari, 2016).

Longitude dalam bahasa Indonesia adalah garis lintang. Garis lintang merupakan garis yang biasa digunakan dalam menentukan suatu lokasi yang berpusat pada garis khatulistiwa. Garis lintang mengitari bumi dari bagian ekuator hingga pada bagian kutub utara dan kutub selatan. Garis yang berada di sebelah selatan ekuator dikenal dengan istilah Lintang Selatan (LS), sedangkan garis yang berada di sebelah utara ekuator dikenal dengan istilah Lintang Utara (LS). Jarak antara garis satu dengan yang lainnya dihitung dengan satuan derajat. Dalam ilmu geografi, satuan derajat juga disebut sebagai jam. 1 yang terbagi menjadi 60 menit dan 1 menit terbagi menjadi 60 detik. Sehingga dalam hal ini, garis lintang pada suatu lokasi dapat juga disebut sebagai jam (Sari, 2016).

Brightness dalam bahasa Indonesia adalah kecerahan atau dengan dengan makna lain adalah tingkat kecerahan titik api yang ada pada suatu wilayah. Variabel *brightness* digunakan

sebagai indikator untuk menentukan seberapa parah kejadian titik api pada suatu wilayah.

Acq_date merupakan variabel dalam data kejadian titik api yang menjabarkan data tanggal, bulan, dan tahun kejadian titik api dalam studi kasus ini. *Acq_date* dapat digunakan untuk menemukan pola waktu kejadian titik api, sehingga variabel *acq_date* diasumsikan sebagai variabel yang penting untuk di klaster.

Convidence dalam bahasa Indonesia adalah keyakinan, dalam istilah adalah tingkat keyakinan untuk mengukur suatu kenyataan atau kejadian. Dalam studi kasus ini *convidence* digunakan sebagai indikator untuk mengukur tingkat keyakinan kejadian titik api di beberapa wilayah.

2.2.1. Algoritme K-Means

K-Means merupakan salah satu algoritme klasterisasi yang dapat membagi data ke dalam beberapa klaster. Algoritme *K-Means* dapat mempartisi data pada kelompok sehingga data yang memiliki karakteristik sama dikumpulkan dalam satu kelompok atau yang lebih dikenal dengan klaster. Sebaliknya data yang memiliki karakteristik yang berbeda akan dimasukkan ke dalam kelompok yang lain. Dalam hal ini, tujuan dari pengelompokan data adalah untuk meminimalisasi fungsi objektif yang diatur dalam proses pengelompokan (Agustin, et al., 2015).

Pada dasarnya algoritme *K-Means* melakukan dua proses yaitu pendeteksian lokasi pusat klaster dan proses pencarian anggota dari tiap-tiap klaster. Berikut merupakan proses dari algoritma *K-Means*:

1. Menentukan *K* sebagai jumlah klaster yang akan dibentuk dan menetapkan pusat klaster
2. Mengalokasikan data kedalam klaster secara acak
3. Menghitung pusat klaster dengan menggunakan Persamaan (1).

$$C_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p} \quad (1)$$

Keterangan:

C_{kj} = Pusat klaster ke *k* pada fitur ke *j*

x_{ij} = Fitur *j* pada data ke *i*

p = Jumlah data

4. Menghitung jarak setiap data ke pusat klaster dengan Persamaan (2).

$$d_{ik} = \sqrt{\sum_j^m (x_{ij} - C_{kj})^2} \quad (2)$$

Keterangan:

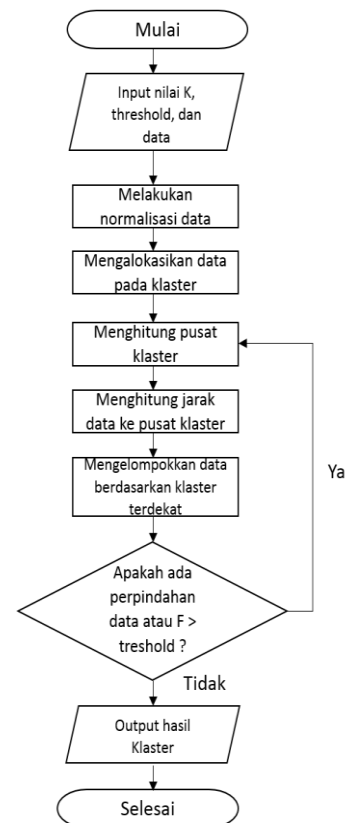
d_{ik} = Jarak data *i* ke pusat klaster *k*

m = Jumlah fitur

x_{ij} = Fitur *j* pada data ke *i*

C_{kj} = Fitur *j* pada Pusat klaster *k*

Jika terdapat data yang berpindah klaster atau, atau jika perubahan nilai pada fungsi objektif masih diatas nilai ambang maka dilakukan iterasi dimulai dari langkah tiga. Gambar 3 menggambarkan diagram alir algoritme *K-Means*.



Gambar 3. Diagram alir algoritme *K-Means*

2.2.2. Algoritme Isodata

Algoritme *Isodata* merupakan salah satu algoritme klasterisasi yang memungkinkan jumlah klaster disesuaikan secara otomatis berdasarkan proses iterasi yang dilakukan. Berbeda halnya dengan algoritme *K-Means*, Algoritme *Isodata* menentukan jumlah klaster pada tahap awal dan jumlah klaster tidak dapat berubah sepanjang iterasi yang dilakukan. Algoritme *Isodata* bekerja dengan menggabungkan klaster yang sama dan data

pada suatu kluster dapat membelah berdasarkan simpangan baku yang besar (Wang, 2016). Berikut merupakan parameter yang digunakan pada Algoritme *Isodata*.

K_0 = Jumlah kluster awal
 n_{min} = Jumlah data minimum kluster
 σ^2_{max} = Varian maksimum (*merge*)
 d_{min} = Jarak berpasangan minimum (*split*)

Berikut merupakan tahapan algoritme *Isodata*:

1. Memilih secara acak nilai $K = K_0$ dan menghitung nilai rata rata vektor $\{m_1, \dots, m_k\}$ dari data set.

2. Menentukan titik data x untuk kluster dengan nilai rata rata terdekat menggunakan Persamaan (3).

$$x \in \omega_i \text{ if } d(x, m_i) = \min\{d(x, m_1), \dots, d(x, m_k)\} \quad (3)$$

Keterangan:

x = data

m = Pusat Kluster

3. Menghilangkan kluster yang hanya memiliki sedikit anggota yaitu jika $n_j < n_{min}$ maka buang kluster dan memindahkan data pada kluster lain.

4. Melakukan pembaruan nilai rata rata vektor pada kluster $\omega_j (j = 1, \dots, K)$ dengan Persamaan (4).

$$m_j = \frac{1}{n_j} \sum_{x \in \omega_j} x, \quad (4)$$

5. Jika $K \leq K_0 / 2$, maka lakukan langkah 6 untuk melakukan pemisahan kluster dan jika $K > 2K_0$ maka lakukan langkah 7 untuk melakukan penggabungan kluster.

6. Membagi setiap kluster $\omega_j (j = 1, \dots, K)$, dan menghitung kovarian terbesar dengan Persamaan (5).

$$\sigma^2_m = \max\{\sigma^2_1, \dots, \sigma^2_N\}, \text{ Jika } \sigma^2_m > \sigma^2_{max} \text{ dan } n_j > 2n_{min}, \quad (5)$$

maka membagi dua pusat kluster baru,

$$m_j^+ = m_j + \sigma_m, \quad m_j^- = m_j - \sigma_m,$$

Atau menggunakan PCA untuk mencari varian sesuai dengan *eigen value* terbesar

λ_{max} dan membagi kluster berdasarkan vektor *eigen*.

7. Menggabungkan kluster $K(K-1)/2$ dengan menggunakan *Euclidean distance*. Penggabungan kluster, dapat menggunakan metode yang berbeda. Apa bila hasil d_B memenuhi kondisi $d_B(\omega_i, \omega_j) < d_{min}$, maka kedua kluster tersebut, digabungkan dengan Persamaan (6).

$$m_i = \frac{1}{n_i + n_j} [n_i m_i + n_j m_j] \quad (6)$$

hapus m_j , set $K \leftarrow K - 1$

8. Kembali pada langkah 2 apabila jumlah iterasi maksimum belum tercapai, Jika iterasi maksimum tercapai, maka berhenti melakukan iterasi.

Gambar 4 menggambarkan diagram alir algoritme *isodata*

2.3. Silhouette Coefficient (SC)

Silhouette Coefficient (SC) merupakan sebuah metode untuk menghitung nilai objek yang berada pada suatu kluster. Metode ini merupakan metode gabungan dari *cohesion* dan *separation* (Wahyuni, et al., 2016). Berikut merupakan proses yang dilakukan dalam metode *Silhouette Coefficient*.

1. Menghitung rata-rata jarak antar objek yang berada dalam satu kluster yang sama, dengan Persamaan (7).

$$a(i) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \quad (7)$$

Keterangan:

j = objek lain dalam suatu kluster

$d(i, j)$ = jarak antar objek i dengan j

2. Menghitung jarak antar objek yang berbeda kluster dan mengambil nilai yang terkecil pada Persamaan (8).

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \quad (8)$$

Keterangan:

$d(i, C)$ = jarak rata-rata objek i dengan objek C dimana,
 $A \neq C$

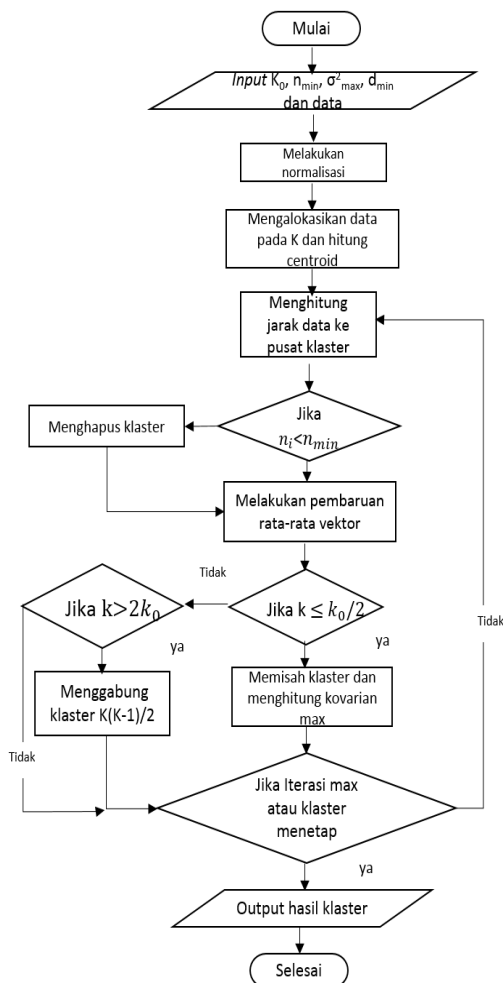
$$b(i) = \min_{C \neq A} d(i, C)$$

3. Menghitung nilai *Silhouette Coefficient* dengan menggunakan Persamaan (9).

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (9)$$

Keterangan:

$s(i)$ = hasil akhir penentuan nilai SC.



Gambar 4. Diagram alir algoritme *Isodata*

2.4. Pengujian dan analisis

Pengujian dan analisis merupakan tahapan untuk melakukan evaluasi hasil dari implementasi yang telah dilakukan. Dalam penelitian ini, hasil kluster dari penerapan algoritme *K-Means* dan *Isodata* di uji berdasarkan kualitas kluster yang dihasilkan. Metode yang digunakan untuk mengukur kualitas kluster dalam penelitian ini, adalah *Silhouette Coefficient* (SC). Hasil dari pengujian yang dilakukan dengan metode *Silhouette Coefficient* akan di jadikan sebagai acuan dalam melakukan analisis perbandingan antara metode *K-Means* dan *Isodata* pada studi kasus

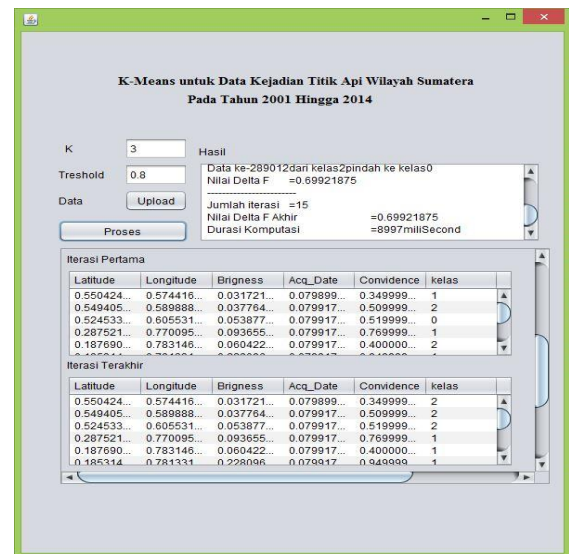
klusterisasi data kejadian titik api di wilayah sumatera pada tahun 2001 hingga 2014.

3. IMPLEMENTASI

Implementasi dilakukan dengan cara menerapkan hasil perancangan pada sub bab 2.2. Berikut merupakan implementasi algoritme *K-Means* dan *Isodata*:

3.1. Implementasi *K-Means*

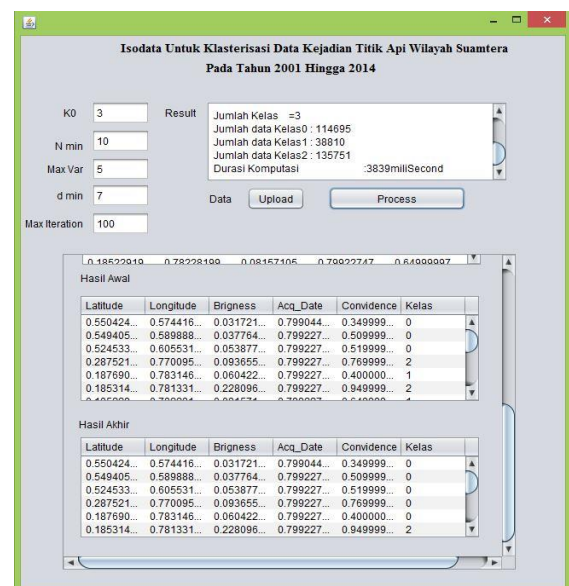
Gambar 5 menampilkan antarmuka implementasi algoritme *K-Means*.



Gambar 5. Implementasi algoritme *K-Means*

3.1. Implementasi *Isodata*

Gambar 6 menampilkan antarmuka implementasi algoritme *Isodata*.



Gambar 6. Implementasi algoritme *Isodata*

4. PENGUJIAN DAN ANALISIS

Dalam penelitian ini, pengujian dilakukan dengan melakukan uji coba pada masukan variabel yang ada pada program *K-Means* dan *Isodata*. Masukan variabel yang diuji, diukur berdasarkan hasil nilai *Shilhouette Coefficient*. Dalam hal ini bahasan pengujian yang diangkat adalah pengaruh variabel masukan algoritme *K-Means* dan *Isodata* terhadap nilai SC dan analisis perbandingan kedua metode tersebut. Data yang digunakan pada uji coba, merupakan data sampel sebanyak 10 ribu data dari 289.257 data kejadian titik api.

4.1. Uji coba *K-Means*

Pada algoritme *K-Means* variabel yang di uji adalah K dan *threshold*. Variabel K dan *threshold* diuji dengan beberapa nilai yang memungkinkan untuk dijadikan sebagai nilai masukan variabel. Dari beberapa nilai masukan akan dipilih salah satu untuk dijadikan variabel masukan pada uji coba pada tahap selanjutnya. Nilai yang terpilih adalah nilai yang memiliki nilai rata-rata SC tertinggi. Pada variabel K nilai SC tertinggi sebesar 0.9999972 untuk $K = 4$ dan variabel *threshold* sebesar 0.999997373 untuk *threshold* = 0.3.

4.2. Uji coba *Isodata*

Pada algoritme *Isodata* variabel yang di uji adalah K_0 , n_{min} , σ^2_{max} , d_{min} dan iterasi Maksimal. Variabel tersebut diuji dengan beberapa nilai yang memungkinkan untuk dijadikan sebagai nilai masukan variabel. Dari beberapa nilai masukan akan dipilih salah satu untuk dijadikan variabel masukan pada uji coba pada tahap selanjutnya. Nilai yang terpilih adalah nilai yang memiliki nilai rata-rata SC tertinggi. Pada variabel nilai SC tertinggi sebesar 0.99999024 untuk $K_0 = 7$, variabel n_{min} sebesar 0.999997315 untuk $n_{min} = 1000$, variabel σ^2_{max} sebesar 0.999997904 untuk $\sigma^2_{max} = 0.6$, variabel d_{min} sebesar 0.99999791, untuk $d_{min} = 0.2$ dan variabel iterasi maksimal sebesar 0.9999976, untuk *iterasi max* = 30.

4.2. Analisis perbandingan

Perbandingan kualitas kluster merupakan tahapan yang membahas tentang perbandingan metode *K-Means* dan *Isodata* berdasarkan nilai dari kualitas kluster yang dihasilkan dengan metode *Silhouette Coefficient* (SC). Perbandingan kualitas kluster dilakukan dengan cara

memasukkan nilai variabel terbaik dengan pengujian 10.000 data yang telah dilakukan. Seluruh variabel terbaik pada metode *K-Means* dan *Isodata* di input pada masing-masing program dan dilakukan uji coba sebanyak lima kali percobaan. Nilai akhir yang dijadikan sebagai tolak ukur kualitas kluster kedua metode, merupakan nilai rata-rata SC dari lima kali percobaan yang dilakukan. Pada algoritme *K-Means* nilai variabel yang dimasukkan adalah $K = 4$ dan *threshold* = 0.3. Pada algoritme *Isodata* nilai variabel yang dimasukkan adalah $K_0 = 7$, $n_{min} = 1000$, $\sigma^2_{max} = 0.6$, $d_{min} = 0.2$ dan *Iterasi Max* = 30. Berdasarkan hasil nilai rata-rata yang didapat pada Tabel 5.8, Metode *K-Means* menghasilkan nilai kluster yang lebih tinggi dibandingkan dengan *Isodata* dengan nilai 0.999997187. Hasil yang diperoleh dapat dipengaruhi oleh variasi nilai K dan *threshold* pada Algoritme *K-Means*. Sedangkan pada algoritme *Isodata* hasil yang diperoleh dapat dipengaruhi oleh variasi nilai K_0 , n_{min} , σ^2_{max} , d_{min} dan *Iterasi Max*. Apabila diamati dari segi jumlah variabel algoritme *Isodata* memiliki jumlah variabel yang lebih banyak dari pada algoritme *K-Means*. Dari sudut pandang yang lain, apabila diamati dari segi tahapan algoritme, algoritme *Isodata* memiliki tahapan yang lebih komplek dari pada algoritme *K-Means*. Dalam hal ini algoritme *K-Means* merupakan algoritme yang lebih sederhana dibandingkan algoritme *Isodata*. Kompleksitas yang tidak terlalu dibutuhkan pada algoritme *Isodata* dapat menjadi faktor yang dapat mempengaruhi kualitas kluster algoritme *K-Means* lebih unggul dibandingkan algoritme *Isodata* pada uji coba klusterisasi data kejadian titik api untuk 1000 data.

5. KESIMPULAN

Berdasarkan studi pustaka, pengumpulan data, analisis kebutuhan, perancangan sistem, implementasi, pengujian dan analisis penelitian yang dilakukan dalam penelitian ini maka dapat di Tarik beberapa kesimpulan, yaitu:

1. Algoritme *K-Means* menghasilkan kualitas kluster yang sangat baik berdasarkan nilai SC yang diperoleh. Sehingga metode *K-Means* sangat layak untuk digunakan dalam mengkluster data kejadian titik api di wilayah Sumatera pada tahun 2001 hingga 2014.

2. Algoritme *Isodata* menghasilkan kualitas kluster yang sangat baik berdasarkan nilai SC yang diperoleh. Sehingga metode *Isodata* sangat layak untuk digunakan dalam mengkluster data kejadian titik api di wilayah Sumatera pada tahun 2001 hingga 2014.
3. Berdasarkan hasil pengujian dan analisis yang dilakukan dalam penelitian ini, algoritme *K-Means* menghasilkan nilai kualitas kluster yang lebih tinggi dibandingkan dengan algoritme *Isodata* untuk pengujian 10.000 data. Pada algoritme *K-Means* nilai SC yang diperoleh sebesar 0.999997187 dan algoritme *Isodata* sebesar 0.999957161. Sehingga dalam hal ini, antara metode *K-Means* dan *Isodata* memiliki selisih nilai SC yang kecil.

6. DAFTAR PUSTAKA

- Agustin, F. E. M., Fitria, A. & Hanifah, A., 2015. Implementasi Algoritma K-Means untuk Menentukan Kelompok. *Jurnal Teknik Informatika*, VIII(1), pp. 73-78.
- Ghosh, S. & Dubey, S. K., 2013. Comparative Analysis of K-Means and Fuzzy C-Means Algorithms. *International Journal of Advanced Computer Science and Applications*, IV(4), pp. 35-39.
- Harrison, M. E., Page, S. E. & Limin, S. H., 2009. The Global Impact of Indonesian Forest. *Journal of Biological Education*, LVI(3), pp. 156-163.
- Liu, Q., Zhao, Z., Li, Y.-X. & Li, Y., 2012. Feature selection based on sensitivity analysis of fuzzy ISODATA. *Neurocomputing*, Issue 87, pp. 29-37.
- Muningsih, E. & Kiswati, S., 2015. Penerapan Metode K-Means Untuk Clustering Produk Online Shop. *Bianglala Informatika*, III(1), pp. 10-17.
- Rasyid, F., 2014. Permasalahan dan Dampak Kebakaran Hutan. *Jurnal Lingkar Widyaiswara*, Issue 1, pp. 47-59.
- Sari, M., 2016. *Ilmu Bumi : Garis Bujur*. [Online] Available at: <https://ilmugeografi.com/ilmu-bumi/garis-bujur> [Accessed 25 Oktober 2017].
- Sari, M., 2016. *Ilmu Bumi : Garis Lintang*. [Online] Available at: <https://ilmugeografi.com/ilmu-bumi/garis-lintang> [Accessed 25 Oktober 2017].
- Wahyuni, I., Yudha, A. A., Asyrofa, R. & Wayan, F. M., 2016. Clustering Nasabah Bank Berdasarkan Tingkat Likuiditas Menggunakan Hybrid. *Jurnal Ilmiah Teknologi dan Informasi ASIA (JITIKA)*, X(2), pp. 24-33.
- Wang, R., 2016. *Classification : Unsupervised Classification-Clustering*. [Online] Available at: <http://fourier.eng.hmc.edu/e161/lectures/classification/node12.html> [Accessed 6 Juni 2016].
- Yu, J., Zhou, Z., Zhong, H. & Huang, H., 2012. An Improved Fuzzy ISODATA Algorithm for Credit Risk Assessment of the EIT Enterprises. *Modern Economy*, III(1), pp. 686-689.