

Naskah
Publikasi_DesyExasanti_170155
0048_09062021
by Any Mariawati

Submission date: 09-Jun-2021 09:02AM (UTC+0700)

Submission ID: 1603164122

File name: JurnalPublikasi_DesyExasanti_1701550048.pdf (408K)

Word count: 5301

Character count: 30888

Analisa Hasil Pengelompokan Wilayah Kejadian Non-Kebakaran Menggunakan *Agglomerative Hierarchical Clustering* di Semarang

Desy Exasanti¹, Arief Jananto²

^{1,2}Fakultas Teknologi Informasi, Sistem Informasi, Universitas Stikubank, Semarang, Indonesia

Jl. Tri Lomba Juang No 1 Semarang 50241, Kota Semarang, Jawa Tengah

Email: ¹dexasanti@gmail.com ²ajananto09@edu.unisbank.ac.id

Abstrak—Klasterisasi merupakan metode pengelompokan dari data yang sudah diketahui label kelasnya untuk menemukan klaster baru dari hasil observasi. Dalam klasterisasi banyak metode yaitu metode terpusat, hirarki, kepadatan dan berbasis kisi, namun dalam penelitian yang dilakukan ini dipilih metode berbasis hirarki. Metode hirarki ini bekerja melakukan pengelompokan objek dengan membentuk hirarki klaster namun bukannya berarti selalu digambarkan dengan hirarki dalam organisasi. Dipilihnya *Agglomerative Hierarchical Clustering* dimana merupakan jenis dari bawah ke atas atau biasa disebut (*bottom-up*) dalam metode ini objek yang akan diuji dianggap sebagai objek tunggal sebagai klaster dan lalu dilakukan iterasi untuk menemukan klaster-klaster yang lebih besar. Data yang akan digunakan adalah data non-kebakaran pada Dinas Pemadam Kebakaran Kota Semarang yang mana akan dilakukan pengelompokan wilayah penanganan non-kebakaran. Dinas Pemadam Kebakaran melakukan penanganan bukan hanya kebakaran saja namun ada banyak hal yang sebenarnya dapat ditangani oleh petugas pemadam kebakaran, kejadian non-kebakaran ada beberapa seperti evakuasi reptil, evakuasi kucing, penyelamatan korban kecelakaan dan lain sebagainya. Dari data non-kebakaran dari 16 kecamatan di Kota Semarang pada tahun 2019 akan dilakukan uji menggunakan tiga algoritma yaitu *Single Linkage*, *Average Linkage* dan *Complete Linkage*. Adapun dari algoritma *Single Linkage* dilakukan prosedur pemusatan dari jarak terkecil antar objek data, algoritma *Average Linkage* dilakukan prosedur dari jarak rata-rata objek data, sedangkan jika algoritma *Complete Linkage* dilakukan prosedur pemusatan dari jarak yang terbesar. Implementasi dan visualisasi dari data uji coba yang dilakukan di penelitian ini menggunakan tools WEKA 3.8.4, *Wakaito Environment Analysis for Knowledge* atau yang biasa dikenal dengan WEKA ini merupakan software yang menggunakan bahasa pemrograman java. Dari dataset 380 data diambil sampel 100 data untuk diuji menggunakan WEKA menggunakan metode perhitungan jarak *Manhattan Distance* dengan 3 cluster. Hasil dari data uji coba dapat divisualisasikan dengan visualisasi dendogram pada fitur *visualize tree* dan jika dilakukan visualisasi dalam bentuk grafik dapat dilakukan menggunakan fitur *visualize clusters assignment*.

Kata Kunci: Data Mining, Clustering, WEKA, *Agglomerative Hierarchical Clustering* dan Pengelompokan

Abstract—Clustering is a method of grouping data with known class labels to find new clusters from observations. There are many methods of clustering, namely centralized, hierarchical, density and lattice-based methods, but in this research, a hierarchical-based method was chosen. This hierarchical method works by grouping objects by forming a cluster hierarchy, but the opening means that it is always described by a hierarchy within the organization. *Agglomerative Hierarchical Clustering* is chosen which is a type from the bottom up or commonly called (*bottom-up*) in this method the object to be tested is considered a single object as a cluster and then iterated to find larger clusters. The data that will be used is non-fire data at the Semarang City Fire Department, where the grouping of non-fire handling areas will be carried out. The Fire Department handles not only fires, but there are many things that can actually be handled by firefighters, there are several non-fire incidents such as reptile evacuation, cat evacuation, accident victim rescue and so on. From non-fire data from 16 sub-districts in Semarang City in 2019 a test will be carried out using three algorithms, namely *Single Linkage*, *Average Linkage* and *Complete Linkage*. As for the *Single Linkage* algorithm, the procedure is performed from the smallest distance between data objects, the *Average Linkage* algorithm is performed from the average distance of the data object, whereas if the *Complete Linkage* algorithm is performed the procedure is centralized from the largest distance. The implementation and visualization of the test data carried out in this research uses the tools of WEKA 3.8.4, *Wakaito Environment Analysis for Knowledge* or commonly known as WEKA is software that uses the Java programming language. From the dataset of 380 data, 100 samples were taken to be tested using WEKA using the *Manhattan Distance* calculation method with 3 clusters. The results of the trial data can be visualized with a dendogram visualization on the *visualize tree* feature and if it is visualized in the form of a graph it can be done using the *visualize clusters assignment* feature.

Keywords: Data Mining, Clustering, WEKA, *Agglomerative Hierarchical Clustering* dan Grouping.

1. PENDAHULUAN

1.1. Latar Belakang Masalah

Perkembangan teknologi di Indonesia kini mengalami perkembangan yang begitu pesat, sehingga suatu hal apapun dalam sektor kehidupan itu dapat dimudahkan dengan teknologi. Segala bentuk akses diberbagai

bidang baik pemerintah, kesehatan, pendidikan dapat dengan mudah dan cepat dengan perkembangan teknologi informasi. Adanya sistem informasi dimana perpaduan dari teknologi informasi dan aktivitas pengguna dalam pendukung informasi dapat mendukung dalam suatu instansi untuk melakukan pengolahan data yang ada dalam instansi menjadi informasi yang berguna. Selain itu dapat pula mengembangka rencana yang efektif kedepannya serta meningkatkan produktivitas dalam pengembangan sistem yang ada.

Dinas Pemadam Kebakaran Kota Semarang Jawa Tengah suatu instansi pemerintahan yang bergerak dibidang penyelamatan kebakaran, namun bukan berarti hanya menangani insiden kebakaran saja. Selain itu juga menangani seperti penyemprotan tumpahan oli, *animal rescue* (penyelamatan hewan), penyelamatan ketinggian dan lain sebagainya. Dilihat dari banyaknya yang ditangani sehingga pastinya terdapat penumpukan data di Dinas Pemadam Kebakaran Kota Semarang, namun data yang ada sudah ada dalam bentuk tabel excel. Namun data yang ada hanya masih seputar itu saja adapun visualisasi hanya untuk internal dan nantinya di akhir tahun digunakan untuk evaluasi bersama untuk meningkatkan dalam penanganan kejadian. Selain hal-hal tersebut data yang ada belum dikelola untuk mendapatkan informasi yang bisa berguna kedepannya.

Data mining digunakan untuk setiap sekumpulan data yang ada dapat memeberikan pengetahuan yang penting dan dapat menjadi sesuatu yang bermanfaat bagi organisasi . [1] Dalam penelitian yang dilakukan ini, peneliti tertarik untuk mengelompokkan wilayah berdasarkan waktu kejadian menggunakan *Agglomerative Hierarchical Clustering* (AHC). Adapun untuk *tools* yang digunakan agar lebih mudah dalam mengimplementasikan dan memberikan gambaran yang jelas dari hasil penelitian, peneliti menggunakan *Wakaito Environment for Knowledge Analysis* atau biasa disebut WEKA. Maka dari itu judul penelitian ini adalah Analisa Hasil Pengelompokan Wilayah Kejadian Non-Kebakaran Menggunakan *Agglomerative Hierarchical Clustering* di Semarang Perumusan Masalah

Perumusan Masalah dalam penelitian ini adalah melakukan analisa data pada Dinas Pemadam Kebakaran Kota Semarang untuk menentukan pengelompokan wilayah dengan penanganan kategori non-kebakaran berdasarkan waktu kejadian menggunakan metode *Agglomerative Hierarchical Clustering*.

1.2. Tujuan dan Manfaat Masalah

Adapun tujuan dari penelitian yang dilakukan adalah untuk menghasilkan kluster dari wilayah-wilayah penanganan non-kebakaran di Kota Semarang berdasarkan waktu kejadian berdasarkan data non kebakaran pada Dinas Pemadam Kebakaran Kota Semarang Jawa Tengah.

Manfaat yang didapatkan dari penelitian ini adalah untuk mengetahui pengelompokan wilayah penanganan non-kebakaran waktu kejadian sehingga didapatkan informasi untuk instansi terkait.

1.3. Landasan Teori

a. Data Mining

Data mining merupakan definisi dari proses dimana menemukan dan melakukan deskripsi dari pola didalam data [2], pola yang ada digali untuk didapatkannya informasi yang bermanfaat dan dapat berguna seperti prediksi dikemudian hari.

b. Clustering

Clustering atau klusterisasi merupakan teknik pengelompokan data yang sudah diketahui label kelasnya. *Clustering* bekerja dengan melakukan pengelompokan obyek-obyek data (pola, entitas, kejadian, unit, hasil, dan hasil observasi) kedalam sejumlah *cluster* tertentu. Dengan kata lain teknik ini melakukan segmentasi data ke dalam sejumlah kelompok menurut karakteristik tertentu.[3]

c. Agglomerative Hierarchical Clustering (AHC)

Pengelompokan hirarki (*Hierarchical Clustering*) adalah metode dengan cara melakukan pengelompokan objek-objek data kedalam sebuah kelompok hirarki. [2] *Hierarchical Clustering* dapat dilakukan menggunakan dua strategi, yaitu *bottom-up* yang disebut *agglomerative hierarchical clustering* dan *top-down* yang bisa disebut *divisive hierarchical clustering*. [4]

Agglomerative Hierarchical Clustering, strategi ini dimulai dengan menganggap setiap objek tunggal yang ada sebagai sebuah kluster kemudian secara berulang menggabungkannya untuk membentuk kluster-kluster yang lebih besar. Berikut beberapa algoritma yang ada pada strategi ini :

1. Single Linkage

Algoritma *single linkage* merupakan prosedur pengelompokan pemusatan berdasarkan jarak terkecil antar objek. Algoritma pengelompokan ini diawali dengan memilih jarak terkecil dalam matriks $D = \{d_{ij}\}$, kemudian menggabungkan objek yang bersesuaian misalnya U dan V untuk

memperoleh cluster (UV). Langkah berikutnya adalah mencari jarak antara (UV) dengan cluster lainnya, misalnya W sehingga dapat dituliskan sebagai berikut:

$$d_{(UV)W} = \min(d_{UW}, d_{VW}) \quad (1)$$

d_{UW} adalah jarak tetangga terdekat dari cluster U dan W serta d_{VW} adalah jarak tetangga terdekat dari cluster V dan W. [5]

2. Complete Linkage

Algoritma *complete linkage* merupakan prosedur pengelompokan agglomerative berdasarkan jarak terbesar antar objek. Algoritma ini diawali dengan memilih jarak terbesar dalam matriks $D = \{d_{ij}\}$, kemudian menggabungkan objek yang bersesuaian misalnya U dan V untuk memperoleh cluster (UV). Langkah berikutnya adalah mencari jarak antara (UV) dengan cluster lainnya, misalnya W sehingga dapat dituliskan sebagai berikut:

$$d_{(UV)W} = \max(d_{UW}, d_{VW}) \quad (2)$$

d_{UW} adalah jarak tetangga terjauh dari cluster U dan W serta d_{VW} adalah jarak tetangga terjauh dari cluster V dan W. [5]

3. Average Linkage

Algoritma *average linkage* merupakan prosedur pengelompokan pemusatan berdasarkan rata-rata antar objek. Algoritma average linkage diawali dengan mendefinisikan matriks $D = \{d_{ij}\}$ untuk memperoleh objek yang paling dekat, sebagai contoh U dan V, kemudian objek ini digabung ke dalam bentuk cluster (UV) dan selanjutnya jarak antara (UV) dengan cluster lainnya W, sehingga dapat dituliskan sebagai berikut:

$$d_{(UV)W} = \frac{d(uw) + d(vw)}{n_{(uv)nw}} \quad (3)$$

$n_{(uv)}$ adalah banyaknya anggota dalam cluster (UV) dan n_w adalah banyaknya anggota dalam cluster W. [5]

d. Waikato Environment for Knowledge Analysis (WEKA)

WEKA merupakan software yang berisi implementasi dari metode-metode yang ada pada data mining. WEKA dikembangkan oleh Universitas Waikato, Selandia Baru menggunakan bahasa pemrograman Java. [6] Software ini didesain agar dapat dengan cepat untuk menguji dengan metode-metode yang ada menggunakan data yang baru dengan fleksibel. [2]

WEKA menyediakan tiga *interface* grafis untuk mengolah data yaitu *Explorer*, *KnowledgeFlow*, dan *Experimenter*. *Explorer* merupakan antarmuka pada WEKA yang menyediakan akses ke semua fungsi-fungsi melalui menu yang urut dan mudah digunakan, seperti *preprocessing*, *clasifier*, *cluster*, *associate*, dan *select attribute*. *KnowledgeFlow* menggunakan berbagai algoritma data mining dengan memasukkan data oleh *user* dan proses yang dilakukan secara visual. *Interface* yang ketiga adalah *Experimenter* dimana fungsinya adalah membandingkan kinerja algoritma-algoritma klasifikasi yang ada pada WEKA, seperti akurasi, nilai *Kappa*, *Mean absolute error*, waktu eksekusi, dan lain-lain. [6]

2. METODE PENELITIAN

2.1 Objek Penelitian

Objek penelitian yang digunakan pada penelitian ini adalah data non kebakaran di Dinas Pemadam Kebakaran Kota Semarang Jawa Tengah.

2.2. Sumber Data

Data internal yang merupakan data sekunder dari Dinas Pemadam Kebakaran Kota Semarang yang sudah mendapatkan persetujuan dari instansi untuk dilakukan penelitian.

2.3. Metode Pengumpulan Data

a. Metode Observasi

Metode observasi merupakan metode yang dilakukan dengan melakukan pengumpulan data yang kompleks dikarenakan melibatkan banyak faktor didalamnya. [7] Dalam penelitian ini dilakukan metode pengumpulan data yang dilakukan secara langsung ke instansi untuk memperoleh data yang

dibutuhkan secara langsung. Data yang dibutuhkan merupakan *database* khusus yang diperlukan perizinan untuk mengaksesnya.

b. Metode Wawancara

Metode wawancara dilakukan dengan tatap muka dan melakukan tanya jawab secara langsung, di penelitian ini dilakukan secara tidak terstruktur yang hanya menanyakan poin-poin seputar data yang akan dikelola sebagai bahan penelitian. [7]

c. Studi Kepustakaan

Metode studi kepustakaan dimana merupakan kajian teoritis, referensi serta literatur ilmiah yang berkaitan dengan budaya, norma yang berkembang dalam situasi sosial yang diteliti. [8]

2.4. Metode Analisa Data

Metode analisa data yang digunakan adalah *Knowledge Discovery and Data Mining* (KDD) dimana dalam pencarian pengetahuan dalam data dapat ditemukan menggunakan lima proses berurutan yaitu seleksi, pemrosesan data, transformasi data, data mining dan interpretasi[4]. Sebelum memasuki tahapan KDD dapat dilihat data mentah dari Data Non Kebakaran Dinas Pemadam Kebakaran Kota Semarang dilihat pada tabel berikut :

Tabel 1. Data Mentah Non Kebakaran Dinas Pemadam Kebakaran Kota Semarang

NO	Hari / Tanggal	Waktu Non Pemadaman		Armada PMK	Nama Kegiatan	Pelaksanaan Tugas		Peralatan Yang Digunakan	Lokasi
		Jam Awal	Jam Akhir			TON	JML		
1	Kamis 03 Januari 2019	19.00	21.00	1 Armada	Evakuasi Sarang Tawon	2	6	Bambu	Jl. Jati Raya, Sronдол Wetan Banyumanik
2	Jumat 04 Januari 2019	21.30	23.00	1 Armada	Evakuasi Sarang Tawon	1	6	APD, Senter, Helm, Bambu, Kain Goni	Jl. Kalicari 3 RT.07 RW.03, Pedurangan

Penjelasan pada nama atribut tabel data diatas dapat disimak di tabel selanjutnya di Tabel 2 .

Tabel 2. Keterangan Nama Atribut pada Data Mentah Non Kebakaran

Hari/Tanggal	Hari dan tanggal kejadian non kebakaran.
Jam Awal (Waktu Non Pemadam)	Waktu awal petugas melakukan evakuasi kejadian non kebakaran.
Jam Akhir (Waktu Non Pemadam)	Waktu akhir petugas selesai melakukan evakuasi kejadian non kebakaran.
Armada PMK	Jumlah armada pemadam kebakaran yang menuju lokasi kejadian.
Nama Kegiatan	Klasifikasi kegiatan non pemadam kebakaran.
TON (Pelaksanaan Tugas)	Berat barang atau benda yang dievakuasi oleh petugas.
JML (Pelaksanaan Tugas)	Jumlah petugas yang melakukan eksekusi kejadian.
Lokasi	Alamat terjadinya kejadian non kebakaran.

Setelah mengetahui data mentah yang ada dapat dilakukan tahap metode analisa sebagai berikut :

1. Seleksi (*Selection Data*)

Dalam proses ini data diambil dan diseleksi dari data yang ada untuk dapat dianalisa dalam penelitian ini sehingga dapat dilakukan pemrosesan data mining. Adapun hasil *selection data* fapat dilihat pada tabel 3.

Tabel 3. Hasil *Selection Data*

NO	Tanggal	Nama Kegiatan	Jam Awal	Durasi	Kecamatan
----	---------	---------------	----------	--------	-----------

1	03 Januari 2019	Evakuasi Sarang Tawon	19.00	02.00	Banyumanik
2	04 Januari 2019	Evakuasi Sarang Tawon	21.30	01.30	Pedurungan
3	05 Januari 2019	Evakuasi Sarang Tawon	18.30	02.30	Pedurungan
4	05 Januari 2019	Evakuasi Sarang Tawon	21.30	01.30	Genuk
5	06 Januari 2019	Evakuasi Sarang Tawon	14.15	01.30	Semarang Selatan

2. Pemrosesan Data

Pada tahapan kali ini dilakukan pembuangan data duplikasi, data salah cetak atau inkonsisten juga data yang kosong dari data yang ada.

3. Transformasi Data

Dalam tahap ini dilakukan transformasi data dan konversi data yang belum memiliki entitas yang jelas sehingga dapat dikelola dalam data yang siap diproses dengan data mining.

a.) Transformasi

Atribut yang akan ditransformasi dalam penelitian ini adalah ebagai berikut :

1.) Atribut Nama Kegiatan

Data didalam atribut Nama Kegiatan adalah Evakuasi Ular Evakuasi Kucing Evakuasi Biawak Evakuasi Sapi Evakuasi Sarang Tawon Evakuasi Sarang Lebah Evakuasi Iguana Evakuasi Monyet Penyemprotan Tumpahan Oli Penyemprotan Tumpahan Solar Penyemprotan Tumpahan Minyak Goreng Penyemprotan Tumpahan Tepung Penyiraman Lapangan Penyiraman Kantor Evakuasi Mobil Terperosok Pohon Tumbang Evakuasi Orang Kesetrum Pelepasan Cincin Perbaikan Tali Tiang yang Aus. Hasil dari transformasi atribut nama kegiatan akan dijelaskan pada tabel 5.

Tabel 4. Keterangan Transformasi Atribut Nama Kegiatan

No	Nama Kegiatan	Kasifikasi	Transformasi Data
1.	Evakuasi Ular		
2.	Evakuasi Kucing		
3.	Evakuasi Biawak		
4.	Evakuasi Sapi		
5.	Evakuasi Sarang Tawon	Sarang	Penyelamatan Hewan
6.	Evakuasi Sarang Lebah	Sarang	
7.	Evakuasi Iguana		
8.	Evakuasi Monyet		
9.	Penyemprotan Tumpahan Oli		
10.	Penyemprotan Tumpahan Solar		
11.	Penyemprotan Tumpahan Minyak Goreng	Minyak	Penyemprotan
12.	Penyemprotan Tumpahan Tepung		
13.	Penyiraman Lapangan		
14.	Penyiraman Kantor		
15.	Evakuasi Mobil Terperosok	Mobil	
16.	Pohon Tumbang		Lain-lain
17.	Evakuasi Orang Kesetrum	Orang	
18.	Pelepasan Cincin		

19. Perbaikan Tali Tiang yang Aus

2.) Jam Awal

Pada penelitian kali ini diambil atribut untuk diketahuinya waktu kejadian non kebakaran. Adapun waktu dibagi menjadi pagi, siang, sore dan malam dalam kategori waktu pagi dari jam 05.00 WIB– 10.00 WIB, lalu siang jam 10.00 WIB – 15.00 WIB, sore jam 15.00 WIB – 18.00 WIB dan malam 18.00 WIB – 05.00 WIB . Untuk lebih jelasnya dapat dilihat pada hasil transformasi atribut jam awal di tabel 6.

Tabel 5. Tabel Transformasi Atribut Jam Awal

No.	Nama Data pada Atribut	Transformasi Data
1.	05.00 – 10.00	Pagi
2.	10.00 - 15.00	Siang
3.	15.00 -18.00	Sore
4.	18.00 – 05.00	Malam

b. Konversi

Data yang sebelumnya sudah ditransformasi maka kemudian dikonversi agar mudah untuk pemrosesan data mining. Adapun atribut yang dikonversi adalah sebagai berikut :

1.) Nama Kegiatan

Data yang digunakan pada atribut Nama Kegiatan adalah penyelamatan hewan, penyemprotan, dan lain-lain. Selanjutnya akan dikonversi untuk perhitungan dalam proses data mining. Hasil Konversi atribut Nama Kegiatan dapat dilihat pada tabel berikut.

Tabel 6. Tabel Konversi Atribut Nama Kegiatan

Nama Kegiatan	Konversi Data
Penyelamatan Hewan	1
Penyemprotan	2
Lain-lain	3

2.) Jam Awal

Data yang digunakan pada atribut jam awal sangatlah beragam untuk itu dapat ditransformasi kedalam 4 kategori dan lalu dikonversikan agar dapat lebih mudah diproses kedalam data mining. Hasil konversi atribut Jam Awal dapat dilihat ditabel berikut.

Tabel 7. Tabel Konversi Atribut Jam Awal

Kategori Waktu	Konversi Data
Pagi	1
Siang	2
Sore	3
Malam	4

3.) Durasi

Data yang ada pada atribut durasi juga sangatlah beragam untuk itu dapat dikelompokkan dengan > 1 jam, 1-2 jam, 2-3 jam, <3 jam .Lalu dapat dikonversikan dan dilihat hasilnya di tabel 9. agar dapat lebih mudah kedalam proses data mining.

Tabel 8. Tabel Konversi Atribut Durasi

Durasi	Konversi Data
>1 jam	1
1-2 jam	2
2-3 jam	3
<3 jam	4

4.) Kecamatan

Tabel 9. Tabel Konversi Atribut Kecamatan

Kecamatan	Konversi Data
Banyumanik	1
Genuk	2
Pedurungan	3
Tembalang	4
Mijen	5
Ngaliyan	6
Gayamsari	7
Candisari	8
Gajah Mungkur	9
Gunungpati	10
Tugu	11
Semarang Selatan	12
Semarang Barat	13
Semarang Utara	14
Semarang Tengah	15
Semarang Timur	16

Kemudian hasil dari seluruh proses *data preprocessing* dapat dilihat di tabel berikut.

Tabel 10. Hasil Data *Preprocessing*

Nama Kegiatan	Jam Awal	Durasi	Kecamatan
1	4	2	1
1	4	2	3
1	4	3	3
1	4	2	2
1	2	2	12

4. *Data mining*, proses esensial untuk mengekstrak pola dari data dengan metode cerdas. Pada tahap ini digunakan metode *Agglomerative Hierarchical Clustering* (AHC) dan menggunakan algoritma *Single Linkage*, *Average Linkage* dan *Complete Linkage*.
5. Interpretasi atau evaluasi,
 Pada tahapan yang terakhir dilakukanlah proses pembentukan *output* yang dapat mudah dimengerti dan bersumber pada proses Data Mining. Adapun output nantinya dapat dievaluasi apabila adanya kesalahan.

3. HASIL DAN PEMBAHASAN

4.1. Perhitungan Manual Perhitungan Jarak Matrik Data

Ada berbagai macam perhitungan jarak antar objek data, seperti jarak antar atribut nominal, atribut biner, atribut numerik, atribut ordinal dan lain sebagainya. Namun dalam penelitian kali ini data yang digunakan sudah dikonversi sehingga menggunakan perhitungan jarak atribut numerik, dan menggunakan metode *Manhattan Distance* atau juga dikenal dengan *Taxicab Distance* dimana metode ini mengadaptasi jarak antar dua lokasi didalam sebuah kota. Adapun rumus dari *Manhattan Distance* ini sebagai berikut :

$$D(u,v) = |x_{u1} - x_{v1}| + |x_{u2} - x_{v2}| + \dots + |x_{up} - x_{vp}| \quad (4)$$

Dimana $|x_{ui} - x_{vi}|$ menyatakan selisih absolut antara nilai atribut ke-1 pada objek u dan nilai atribut ke-1 pada objek v . [3]

Dalam data yang diolah menggunakan data konversi lalu dihitung semua pasangan dua data menggunakan *Manhattan Distance*

1. Data (1,1)

$$D(1,1) = |1 - 1| + |4 - 4| + |2 - 2| + |1 - 1| \\ = 0$$

2. Data (1,2)

$$D(1,2) = |1 - 1| + |4 - 4| + |2 - 2| + |3 - 1| \\ = 0 + 0 + 0 + 2 = 2$$

3. Data (1,3)

$$D(1,3) = |1 - 1| + |4 - 4| + |3 - 2| + |3 - 1| \\ = 0 + 0 + 1 + 2 = 3$$

4. Data (1,4)

$$D(1,4) = |1 - 1| + |4 - 4| + |2 - 2| + |2 - 1| \\ = 0 + 0 + 0 + 1 = 1$$

5. Data (1,5)

$$D(1,5) = |1 - 1| + |4 - 2| + |2 - 2| + |2 - 1| \\ = 0 + 2 + 0 + 1 = 3$$

6. Data (2,3)

$$D(2,3) = |1 - 1| + |4 - 4| + |3 - 2| + |3 - 3| \\ = 0 + 0 + 1 + 0 = 1$$

7. Data (2,4)

$$D(2,4) = |1 - 1| + |4 - 4| + |2 - 2| + |3 - 2| \\ = 0 + 0 + 0 + 1 = 1$$

8. Data (2,5)

$$D(2,5) = |1 - 1| + |4 - 2| + |2 - 2| + |2 - 3| \\ = 0 + 2 + 0 + 9 = 11$$

9. Data (3,4)

$$D(3,4) = |1 - 1| + |4 + 4| + |3 - 2| + |3 - 2| \\ = 0 + 0 + 1 + 1 = 2$$

10. Data (3,5)

$$D(3,5) = |1 - 1| + |4 - 2| + |3 - 2| + |2 - 3| \\ = 0 + 2 + 1 + 9 = 12$$

11. Data (4,5)

$$D(4,5) = |1 - 1| + |4 - 2| + |2 - 2| + |2 - 2| \\ = 0 + 2 + 0 + 10 = 12$$

Hasil perhitungan matriks menggunakan metode *Manhattan Distance* dapat dilihat pada Tabel 16 berikut.

Tabel 11. Contoh Hasil Perhitungan Jarak Matrik Data

Data	1	2	3	4	5
1	0	2	3	1	13
2	2	0	1	1	11
3	3	1	0	2	12
4	1	1	2	0	12
5	13	11	12	12	0

4.2. Implementasi Menggunakan Software WEKA

Dalam penelitian ini digunakan software WEKA 3.8.4, adapun data yang ada sebanyak 380 *record* dari Data Non Kebakaran Kota Semarang dengan empat atribut, yaitu : Nama Kegiatan, Jam Awal, Durasi dan Kecamatan. Data yang dimasukan rata-rata menggunakan format *Attribute-Relation File Format* (.arff) atau *Comma Separated Values* (.csv). Diambil 100 data sampel serta 3 *cluster* untuk masing-masing algoritma yaitu *Single Linkage*, *Average Linkage* dan *Complete Linkage*. [9]

Hasil uji coba dari masing-masing algoritma dengan menggunakan WEKA 3.8.4 dapat dilihat sebagai berikut.

1. Data sampel sebanyak 100 data dengan dibentuk 3 *cluster* menggunakan metode perhitungan jarak *Manhattan Distance* dan algoritma *Single Linkage* dengan memilih pada *show properties* pada *distanceFunction* dipilih **ManhattanDistance** – *R first – last* dan pastikan pada *numClusters* disesuaikan dengan jumlah *cluster* yang akan dibuat ada pada tabel 17. Berikut.

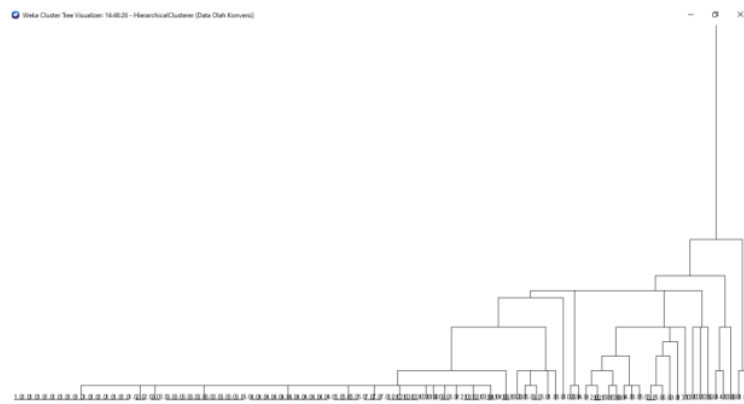
Tabel 12. Hasil Clustering Menggunakan Algoritma Single Linkage

Cluster ke-	Jumlah Anggota Cluster
1	97
2	2
3	1

Data yang telah diuji sebanyak 100 diperoleh 3 *cluster* yaitu dengan *cluster-1* berjumlah 97 (97%), *cluster-2* berjumlah 2 (2%) dan *cluster-3* berjumlah 1 (1%) .

Hasil visualisasi clustering dengan algoritma *Single Linkage* jika disajikan dengan dendrogram dari data 100 sampel dapat dilihat dari gambar 3. Berikut.

Gambar 3. Visualisasi Dendrogram Algoritma Single Linkage



Data yang telah diuji menggunakan algoritma *Single Linkage* menggunakan *visualize clusters assignment* mendapatkan hasil sebagai berikut.

Tabel 13. Potongan Hasil *Visualize Clusters Assignment*

No.	Instance_number	Nama Kegiatan	Jam Awal	Durasi	Kecamatan	Cluster
1	00.00	01.00	04.00	02.00	01.00	cluster0
2	01.00	01.00	04.00	02.00	03.00	cluster0
3	02.00	01.00	04.00	03.00	03.00	cluster0
4	03.00	01.00	04.00	02.00	02.00	cluster0
5	04.00	01.00	02.00	02.00	12.00	cluster0

Hasil analisa data sampel yang telah diuji menggunakan WEKA 3.8.4 didapatkan hasil dengan anggota *cluster* 1 memiliki anggota sebanyak 97. Pada atribut 'Nama Kegiatan' yang mendominasi adalah sebanyak 90 kejadian penyelamatan hewan, pada atribut 'Jam Awal' didominasi dengan waktu kejadian rata-rata terjadi di malam hari dengan sebanyak 78 kejadian dan atribut 'Durasi' didominasi dengan waktu eksekusi kejadian sekitar 1-2 jam dengan banyaknya kejadian sejumlah 77 kejadian. Sedangkan untuk atribut 'Kecamatan' didominasi dengan kecamatan Banyumanik dengan 18 kejadian, Ngaliyan dengan 15 kejadian serta Tembalang dengan 13 kejadian. [10]

Sedangkan untuk *cluster* 2 pada atribut 'Nama Kegiatan' memiliki 2 kejadian dalam kategori lain-lain, pada atribut 'Jam Awal' memiliki 2 kejadian yang terjadi di kategori malam dengan atribut 'Durasi' yang memiliki 2 kejadian yang terjadi dengan kategori eksekusi 1-2 jam. Dan di *cluster* ini di atribut 'Kecamatan' terjadi di Gunung Pati.

Cluster 3 memiliki masing-masing 1 kejadian pada atribut 'Nama Kegiatan' dengan kejadian kategori penyemprotan, 'Jam Awal' dengan kategori sore hari, 'Durasi' dengan kategori 2-3 jam dan 'Kecamatan' di Kecamatan Genuk.

2. Data sampel sebanyak 100 data dengan dibentuk 3 *cluster* menggunakan metode perhitungan jarak *Manhattan Distance* dan algoritma *Average Linkage* dengan memilih pada *show properties* pada *distanceFunction* dipilih *ManhattanDistance - R first - last* dan pastikan pada *numClusters* disesuaikan dengan jumlah *cluster* yang akan dibuat ada pada tabel 19. Berikut.

Tabel 14. Hasil Clustering Menggunakan Algoritma *Average Linkage*

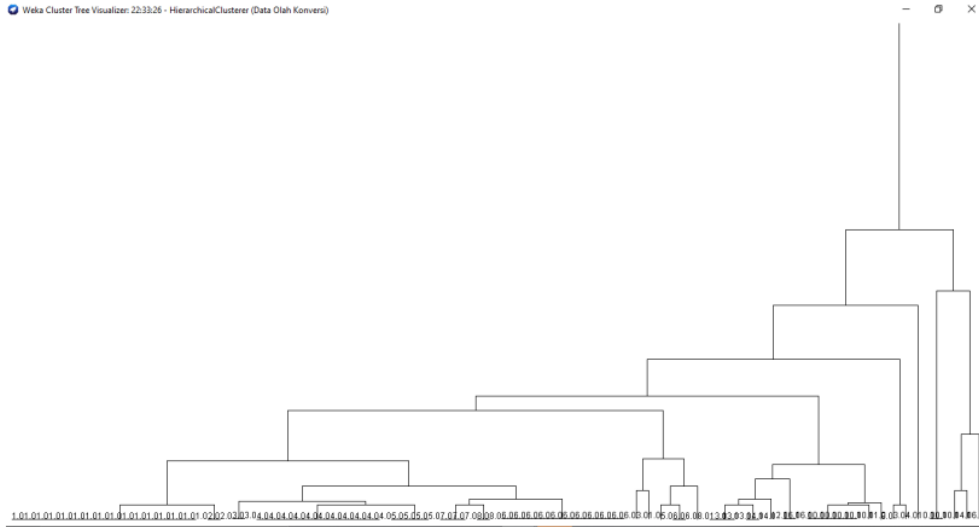
Cluster ke-	Jumlah Anggota Cluster
1	80
2	17
3	3

Data yang telah diuji sebanyak 100 diperoleh 3 *cluster* yaitu dengan *cluster-1* berjumlah 80 (80%), *cluster-2* berjumlah 17 (17%) dan *cluster-3* berjumlah 3 (3%) .

Hasil visualisasi clustering dengan algoritma *Average Linkage* jika disajikan dengan dendrogram dari data 100 sampel dapat dilihat dari gambar 4. Berikut.

Gambar 4. Visualisasi Dendrogram Algoritma *Average Linkage*

Weka Cluster Tree Visualizer: 22:33:26 - HierarchicalClusterer (Data Olah Konversi)



Data yang telah diuji menggunakan algoritma *Average Linkage* menggunakan *visualize clusters assignment* mendapatkan hasil sebagai berikut.

Tabel 15. Potongan Hasil *Visualize Clusters Assignment*

No.	Instance_number	Nama Kegiatan	Jam Awal	Durasi	Kecamatan	Cluster
1	00.00	01.00	04.00	02.00	01.00	cluster0
2	01.00	01.00	04.00	02.00	03.00	cluster0
3	02.00	01.00	04.00	03.00	03.00	cluster0
4	03.00	01.00	04.00	02.00	02.00	cluster0
5	04.00	01.00	02.00	02.00	12.00	cluster1

Hasil analisa data sampel yang telah diuji menggunakan WEKA 3.8.4 didapatkan hasil dengan anggota *cluster 1* memiliki anggota sebanyak 75. Pada atribut 'Nama Kegiatan' yang mendominasi adalah sebanyak 78 kejadian penyelamatan hewan, pada atribut 'Jam Awal' didominasi dengan waktu kejadian rata-rata terjadi di malam hari dengan sebanyak 70 kejadian dan atribut 'Durasi' didominasi dengan waktu eksekusi kejadian sekitar 1-2 jam dengan banyaknya kejadian sejumlah 70 kejadian. Sedangkan untuk atribut 'Kecamatan' didominasi dengan kecamatan Banyumanik dengan 18 kejadian, Ngaliyan dengan 16 kejadian serta Tembalang dengan 15 kejadian. [10]

Sedangkan untuk *cluster 2* pada atribut 'Nama Kegiatan' memiliki 14 kejadian mendominasi dalam kategori penyelamatan hewan, pada atribut 'Jam Awal' memiliki 17 kejadian yang terjadi di kategori pagi dan atribut 'Durasi' yang memiliki 10 kejadian yang terjadi dengan kategori eksekusi 1-2 jam. Dan di *cluster* ini di atribut 'Kecamatan' terjadi 5 kejadian di Mijen.

Cluster 3 memiliki 3 kejadian penyemprotan pada atribut 'Nama Kegiatan', memiliki 1 kejadian di sore hari dan 2 kejadian di malam hari pada atribut 'Jam Awal'. Dalam atribut 'Durasi' memiliki 1 kejadian dengan durasi 2-3 jam dan <3 jam dengan 2 kejadian, sedangkan untuk atribut 'Kecamatan' memiliki masing-masing satu kejadian di Banyumanik, Genuk dan Pedurungan.

3. Data sampel sebanyak 100 data dengan dibentuk 3 *cluster* menggunakan metode perhitungan jarak *Manhattan Distance* dan algoritma *Complete Linkage* dengan memilih pada *show properties* pada *distanceFunction* dipilih *ManhattanDistance - R first - last* dan pastikan pada *numClusters* disesuaikan dengan jumlah *cluster* yang akan dibuat ada pada tabel 21. Berikut.

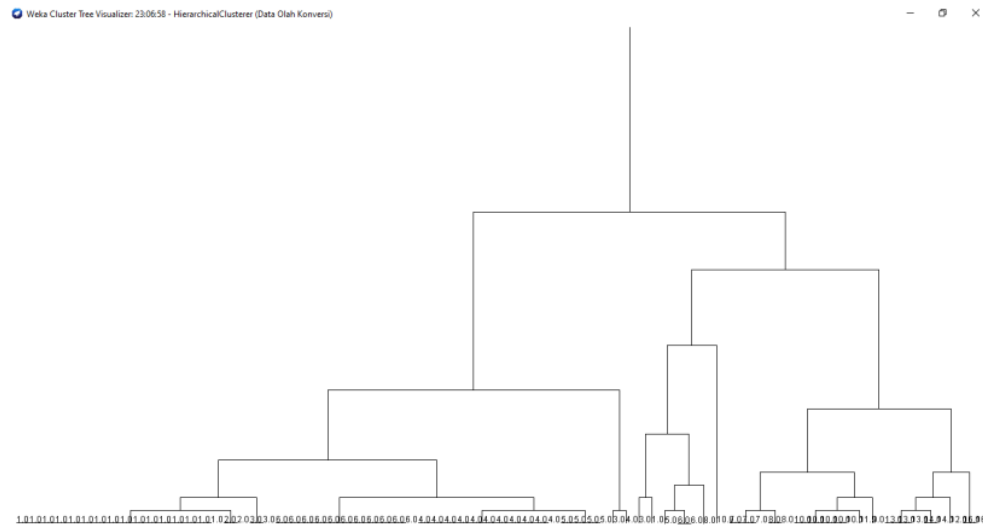
Tabel 16. Hasil Clustering Menggunakan Algoritma *Complete Linkage*

Cluster ke-	Jumlah Anggota Cluster
1	75
2	17
3	8

Data yang telah diuji sebanyak 100 diperoleh 3 cluster yaitu dengan cluster-1 berjumlah 75 (75%), cluster-2 berjumlah 17 (17%) dan cluster-3 berjumlah 8 (8%) .

Hasil visualisasi hasil clustering dengan algoritma *Complete Linkage* jika disajikan dengan dendrogram dari data 100 sampel dapat dilihat dari gambar 5. Berikut.

Gambar 5. Visualisasi Dendrogram Algoritma *Complete Linkage*



Data yang telah diuji menggunakan algoritma *Complete Linkage* menggunakan *visualize clusters assignment* mendapatkan hasil sebagai berikut.

Tabel 17. Potongan Hasil *Visualize Clusters Assignment*

No.	Instance_number	Nama Kegiatan	Jam Awal	Durasi	Kecamatan	Cluster
1	00.00	01.00	04.00	02.00	01.00	cluster0
2	01.00	01.00	04.00	02.00	03.00	cluster0
3	02.00	01.00	04.00	03.00	03.00	cluster0
4	03.00	01.00	04.00	02.00	02.00	cluster0
5	04.00	01.00	02.00	02.00	12.00	cluster1

Hasil analisa data sampel yang telah diuji menggunakan WEKA 3.8.4 didapatkan hasil dengan anggota cluster 1 memiliki anggota sebanyak 75. Pada atribut 'Nama Kegiatan' yang mendominasi adalah sebanyak 72 kejadian penyelamatan hewan, pada atribut 'Jam Awal' didominasi dengan waktu kejadian rata-rata terjadi di malam hari dengan sebanyak 56 kejadian dan atribut 'Durasi' didominasi dengan waktu eksekusi kejadian sekitar

1-2 jam dengan banyaknya kejadian sejumlah 61 kejadian. Sedangkan untuk atribut 'Kecamatan' didominasi dengan kecamatan Banyumanik dengan 18 kejadian, Ngaliyan dengan 15 kejadian serta Tembalang dengan 14 kejadian. [10]

Sedangkan untuk *cluster* 2 pada atribut 'Nama Kegiatan' memiliki 14 kejadian mendominasi dalam kategori penyelamatan hewan, pada atribut 'Jam Awal' memiliki 17 kejadian yang terjadi di kategori siang hari dan atribut 'Durasi' yang memiliki 10 kejadian yang terjadi dengan kategori eksekusi 1-2 jam. Dan di *cluster* ini di atribut 'Kecamatan' terjadi masing-masing 4 kejadian untuk Kecamatan Mijen dan Semarang Barat.

Cluster 3 memiliki 6 kejadian penyempromtsn pada atribut 'Nama Kegiatan', memiliki 7 kejadian di malam hari pada atribut 'Jam Awal'. Sedangkan untuk atribut 'Durasi' memiliki 4 kejadian dengan durasi 2-3 jam dan untuk *cluster* ini memiliki masing-masing 1 kejadian untuk Kecamatan Banyumanik, Genuk, Pedurungan, Gajah Mungkur, Semarang Utara, Semarang Timur dan memiliki 2 kejadian di Kecamatan Gunungpati.

4. KESIMPULAN

4.1. Kesimpulan

Hasil dari penelitian yang dilakukan dengan mengambil objek dari Dinas Pemadam Kebakaran Kota Semarang Jawa Tengah menggunakan *Agglomerative Hierarchical Clustering* (AHC) didapatkan kesimpulan sebagai berikut :

1. Penelitian ini menggunakan algoritma dalam *Agglomerative Hierarchical Clustering* (AHC) dengan menggunakan metode perhitungan jarak yaitu *Manhattan Distance* dan diimplementasikan menggunakan software WEKA atau biasa disebut *Wakaito Environment Analysis for Knowledge* dengan versi 3.8.4 .
2. Digunakannya data sampel 100 dari 380 data non kebakaran tahun 2019 pada Dinas Pemadam Kebakaran Kota Semarang dilakukan uji coba menggunakan WEKA versi 3.8.4 ini menghasilkan yaitu dari algoritma *Single Linkage* diperoleh 3 cluster yaitu dengan cluster-1 berjumlah 97 (97%), cluster-2 berjumlah 2 (2%) dan cluster-3 berjumlah 1 (1%), algoritma *Average Linkage* diperoleh 3 cluster yaitu dengan cluster-1 berjumlah 80 (80%), cluster-2 berjumlah 17 (17%) dan cluster-3 berjumlah 3 (3%) dan algoritma *Complete Linkage* diperoleh 3 cluster yaitu dengan cluster-1 berjumlah 75 (75%), cluster-2 berjumlah 17 (17%) dan cluster-3 berjumlah 8 (8%).
3. Hasil pembagian *cluster* dari hasil uji coba data sampel non kebakaran tersebut memiliki 3 *cluster* dan dapat ditarik beberapa kesimpulan bahwa dalam data non kebakaran yang ditangani oleh Dinas Pemadam Kebakaran Kota Semarang sering terjadi kejadian penyelamatan hewan meliputi evakuasi ular, evakuasi kucing, evakuasi biawak, evakuasi sapi, evakuasi sarang tawon, evakuasi sarang lebah, evakuasi iguana, evakuasi monyet yang dilakukan di malam hari dengan waktu penanganan rata-rata selama 1-2 jam. Sedangkan wilayah di Kota Semarang yang sering dilakukan penanganan non kebakaran yaitu di Kecamatan Banyumanik, Kecamatan Tembalang dan Kecamatan Ngaliyan.
4. Faktor dari tidak banyaknya kejadian selain di 3 kecamatan dengan penanganan paling banyak dapat pula disebabkan kurangnya warga Kota Semarang yang mengetahui informasi bahwa adanya kategori non kebakaran dalam penanganan oleh pemadam kebakaran, sehingga Dinas Pemadam Kebakaran Kota Semarang dapat melakukan sosialisasi terkait kejadian non kebakaran.

4.2. Saran

Dikarenakan belum dirasa sempurna dikarenakan keterbatasan dan kekurangan pada penelitian ini, maka penulis memberikan beberapa saran guna dapat sebagai acuan penelitian lainnya, sebagai berikut :

1. Dalam penelitian kali ini dalam implementasinya dapat menggunakan bahasa pemrograman seperti Rstudio, python dan lain sebagainya.
2. Dapat menggunakan metode klasterisasi selain metode hirarki.
3. Penelitian ini dapat dimanfaatkan oleh Dinas Pemadam Kebakaran Kota Semarang sebagai acuan pengelompokan wilayah penanganan non-kebakaran.

REFERENCES

- [1] M. Bramer. Data for Data Mining Principles of Data Mining . https://doi.org/10.1007/978-1-4471-4884-5_4 .2013 .
- [2] I. H. Witten et al. Data Mining : Practical Machine Learning Tools and Techniques. Edisi Ketiga. Burlington :Morgan Kaufman. 2011.
- [3] Suyanto.Data Mining Untuk Klasifikasi dan Klasterisasi Data. Edisi Revisi. Bandung : Informatika Bandung, 2019.
- [4] J Han et al. 2012. Data Mining : Concepts and Techniques Third. Elsevier.
- [5] Johnson, R.A. & Wichern, D.W., 2002, Applied Multivariate Statistical Analysis, Fifth Edition, New Jersey: Pearson Prentice Inc.
- [6] S. Adinugroho & Y. A. Sari. Implementasi *Data Mining* Menggunakan WEKA.Edisi 1. Malang : UB Press. 2018.
- [7] Samuel. (2016, Agust 16). Judul [Online] Available : <http://ciputrauceo.net/blog/2016/2/18/metode-pengumpulan-data-dalam-penelitian#:~:text=Metode%20pengumpulan%20data%20adalah%20teknik,yang%20digunakan%20untuk%20mengumpulkan%20data>.
- [8] Sugiyono. Metode Penelitian Kuantitatif Kualitatif dan R & D. Bandung : Alfabeta. 2018.
- [9] G. Abdurrahman, “*Clustering* Data Kredit Bank Menggunakan Algoritma *Agglomerative Hierarchical Clustering Average Linkage*”, JUSTINDO, Vol. 4, No.1, p.p. 2502-5724,2019.
- [10] A. N. Fadhilah, “Pengelompokan Minat Baca dan Kebutuhan Literatur Mahasiswa Menggunakan Metode *Agglomerative Hierarchical Clustering* (AHC)”, Sistem Informasi, UNISBANK, Semarang, Indonesia, 2021.

Naskah Publikasi_DesyExasanti_1701550048_09062021

ORIGINALITY REPORT

9%

SIMILARITY INDEX

10%

INTERNET SOURCES

0%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

1

ejurnal.teknokrat.ac.id

Internet Source

4%

2

ejurnal.ung.ac.id

Internet Source

3%

3

core.ac.uk

Internet Source

2%

Exclude quotes On

Exclude matches < 2%

Exclude bibliography On