

**PENERAPAN METODE *AGGLOMERATIVE HIERARCHICAL*
CLUSTERING DALAM PENGELOMPOKAN DATA
PENYAKIT *LEPTOSPIROSIS***



SKRIPSI

**Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Komputer
pada Departemen Ilmu Komputer/ Informatika**

**Disusun oleh:
Awaliza Abdi Sareat
24010312140096**

**DEPARTEMEN ILMU KOMPUTER / INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
2019**

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini:

Nama : Awaliza Abdi Sareat

NIM : 24010312140096

Judul : Penerapan Metode *Agglomerative Hierarchical Clustering* dalam
Pengelompokan Data Penyakit *Leptospirosis*

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang sepengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Semarang, 9 Agustus 2019



Awaliza Abdi Sareat

24010312140096

HALAMAN PENGESAHAN

Judul : Penerapan Metode *Agglomerative Hierarchical Clustering* dalam
Pengelompokan Data Penyakit *Leptospirosis*

Nama : Awaliza Abdi Sareat

NIM : 24010312140096

Telah diujikan pada sidang skripsi dan dinyatakan lulus pada tanggal **9 Agustus
2019**

Semarang, 9 Agustus 2019

Mengetahui,

a. n. Ketua Departemen Ilmu Komputer/

Informatika, FSM UNDIP

sekretaris



Dr. Png. Adi Wibowo, S.Si, M.Kom

NIP. 198203092006041002

Panitia Penguji Skripsi

Ketua,

Dr. Retno Kusumaningrum, S.Si, M.Kom

NIP. 198104202005012001

HALAMAN PENGESAHAN

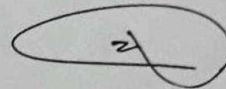
Judul : Penerapan Metode *Agglomerative Hierarchical Clustering* dalam
Pengelompokan Data Penyakit *Leptospirosis*

Nama : Awaliza Abdi Sareat

NIM : 24010312140096

Telah diujikan pada sidang skripsi dan dinyatakan lulus pada tanggal **9 Agustus
2019**

Semarang, 9 Agustus 2019
Pembimbing,



Ragil Saputra, S.Si, M.Cs.

NIP. 198010212005011003

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Penerapan Metode *Agglomerative Hierarchical Clustering* dalam Pengelompokan Data Penyakit *Leptospirosis*”.

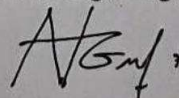
Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana strata satu pada Departemen Ilmu Komputer/Informatika Fakultas Sains dan Matematika Universitas Diponegoro Semarang.

Dalam penyusunan skripsi ini, penulis banyak mendapat bantuan dari berbagai pihak, oleh karena itu penulis ingin mengucapkan rasa hormat dan terima kasih kepada :

1. Ibu Dr. Retno Kusumaningrum, S.Si., M.Kom. selaku Ketua Departemen Ilmu Komputer/Informatika.
2. Bapak Panji Wisnu Wirawan, ST, MT selaku Koordinator Skripsi.
3. Bapak Ragil Saputra, S.Si, M.Cs, selaku dosen pembimbing yang telah memberikan banyak bantuan, pengarahan, waktu, tenaga, pikiran, nasihat, dan saran dalam penyusunan skripsi ini.
4. Bapak Dr. Eng. Adi Wibowo, S.Si, M.Kom, yang telah membantu dalam penentuan topik skripsi ini.
5. Orang tua, keluarga, teman dekat, dan sahabat yang telah mendukung, membantu, dan memberikan semangat kepada penulis dalam menyelesaikan tugas akhir ini.
6. Semua pihak yang telah membantu kelancaran dalam penyusunan tugas akhir, yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan laporan skripsi ini masih banyak kekurangan baik dari segi materi ataupun dalam penyajian. Oleh karena itu kritik dan saran yang membangun sangat penulis harapkan.

Semarang, 9 Agustus 2019



Penulis

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
SKRIPSI UNTUK KEPENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini:

Nama : Awaliza Abdi Sareat
NIM : 24010312140096
Program Studi : Informatika
Departemen : Ilmu Komputer/Informatika
Fakultas : Sains dan Matematika
Jenis karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan **Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty Free Right*)** kepada Universitas Diponegoro atas karya ilmiah saya yang berjudul:

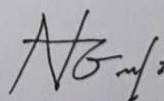
Penerapan Metode Agglomerative Hierarchical Clustering dalam Pengelompokan Data Penyakit Leptospirosis

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-eksklusif ini Universitas Diponegoro berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan skripsi saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/ pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Semarang, 9 Agustus 2019

Yang menyatakan



Awaliza Abdi Sareat

24010312140096

ABSTRAK

Penelitian ini bertujuan untuk menerapkan metode *Agglomerative Hierarchical Clustering* dalam pengelompokan data penyakit *leptospirosis*. *Leptospirosis* adalah penyakit yang disebabkan oleh infeksi bakteri *patogen* yang disebut *leptospira* dan ditularkan dari hewan kepada manusia. Penelitian ini termasuk ke dalam penelitian non-reaktif atau penelitian *unobtrusive*. Sumber data sekunder dalam penelitian ini yaitu rekap data penelitian pasien penderita penyakit *leptospirosis* di Balai Besar Litbang Vektor dan Reservoir Penyakit (B2P2VRP) Salatiga dengan tambahan dari RSUD Sunan Kalijaga Demak tahun 2019 dengan sampel sebanyak 72 data. Penelitian ini menggunakan tiga metode *Agglomerative Hierarchical Clustering* berupa metode *single linkage*, *average linkage* dan *complete linkage* yang masing-masing menggunakan rumus perhitungan *Euclidean Distance*. Hasil penelitian menunjukkan bahwa metode *Agglomerative Hierarchical Clustering single linkage* dalam proses *clustering* data pasien penyakit *leptospirosis* menghasilkan nilai kualitas tertinggi (0.786) jika dibandingkan dengan dua metode lainnya.

Kata Kunci: *Leptospirosis, Agglomerative Hierarchical Clustering, Euclidean Distance*

ABSTRACT

This study aims to apply the Agglomerative Hierarchical Clustering method in grouping data on leptospirosis disease. Leptospirosis is a disease caused by an infection of a pathogenic bacterium called leptospira and transmitted from animals to humans. This is non-reactive research or unobtrusive research. The secondary data sources in this study were data recap of patient research data with leptospirosis disease at Research and Development Center for Disease Vector and Reservoir (B2P2VRP) Salatiga with the addition of data from Sunan Kalijaga Demak Hospital in 2019 with a sample of 72 data. This study uses three Agglomerative Hierarchical Clustering methods, i.e. the single linkage, average linkage and complete linkage, each of which uses the Euclidean Distance calculation formula. The results showed that the single linkage Agglomerative Hierarchical Clustering method in the data clustering process of leptospirosis patients produced the highest quality value (0.786) compared to the other two methods.

Keyword: Leptospirosis, Agglomerative Hierarchical Clustering, Euclidean Distance

DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI	ii
HALAMAN PENGESAHAN	iii
HALAMAN PENGESAHAN	iv
KATA PENGANTAR	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xvii
DAFTAR LAMPIRAN	xviii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan dan Manfaat	3
1.4. Ruang Lingkup	3
1.5. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1. Penelitian Terkait	5
2.2. <i>Leptospirosis</i>	7
2.3. <i>Polymerase Chain Reaction (PCR)</i>	8
2.4. <i>Quantitative Polymerase Chain Reaction (qPCR)</i>	9

2.5.	Normalisasi Data	10
2.6.	<i>Clustering</i>	11
2.7.	<i>Agglomerative Hierarchical Clustering</i>	13
2.7.1.	<i>Single Linkage</i>	15
2.7.2.	<i>Average Linkage</i>	16
2.7.3.	<i>Complete Linkage</i>	16
2.8.	<i>Euclidean Distance</i>	17
2.9.	<i>Davies Bouldin Index</i>	18
BAB III METODOLOGI PENELITIAN		20
3.1.	<i>Data Gathering</i> (Pengumpulan Data).....	20
3.2.	<i>Pre-processing</i> Data (Pengolahan Data Awal)	22
3.2.1.	Transformasi data.....	22
3.2.2.	Normalisasi Data.....	23
3.3.	Perhitungan Metode <i>Agglomerative Hierarchical Clustering</i> dengan <i>Distance Measure</i>	23
3.3.1.	Metode <i>Single Linkage</i> dengan <i>Euclidean Distance</i>	25
3.3.2.	Metode <i>Average Linkage</i> dengan <i>Euclidean Distance</i>	27
3.3.3.	Metode <i>Complete Linkage</i> dengan <i>Euclidean Distance</i>	29
3.4.	Penentuan <i>Cluster</i>	31
3.5.	Evaluasi atau Pengujian	31
BAB IV PEMBAHASAN DAN HASIL		32
4.1.	<i>Data Gathering</i> (Pengumpulan Data).....	32
4.2.	<i>Pre-processing</i> Data (Pengolahan Data Awal)	33
4.2.1.	Transformasi Data.....	33
4.2.2.	Normalisasi Data.....	33

4.3.	Perhitungan Metode <i>Agglomerative Hierarchical Clustering</i> dengan <i>Distance Measure</i>	34
4.3.1.	Metode <i>Single Linkage</i> dengan <i>Euclidean Distance</i>	34
4.3.2.	Metode <i>Average Linkage</i> dengan <i>Euclidean Distance</i>	37
4.3.3.	Metode <i>Complete Linkage</i> dengan <i>Euclidean Distance</i>	41
4.4.	Penentuan <i>Cluster</i>	44
4.4.1.	<i>Single Linkage</i>	44
4.4.2.	<i>Average Linkage</i>	50
4.4.3.	<i>Complete Linkage</i>	56
4.5.	Evaluasi atau Pengujian	62
4.5.1.	<i>Davies Bouldin Index Single Linkage</i>	62
4.5.2.	<i>Davies Bouldin Index Average Linkage</i>	66
4.5.3.	<i>Davies Bouldin Index Complete Linkage</i>	71
4.6.	Pembahasan Hasil <i>Clustering</i>	75
4.7.	Pembahasan <i>Source Code</i>	77
BAB V PENUTUP		83
5.1.	Kesimpulan.....	83
5.2.	Saran.....	83
DAFTAR PUSTAKA		84
LAMPIRAN-LAMPIRAN		88

DAFTAR TABEL

Tabel 2. 1	Daftar Penelitian Terkait.....	5
Tabel 3. 1	<i>Dataset</i> Pasien Penyakit <i>Leptospirosis</i>	21
Tabel 3. 2	<i>Dataset</i> pasien penyakit leptospirosis yang telah diseleksi dan akan digunakan dalam proses <i>Pre-processing</i>	22
Tabel 3. 3	Transformasi Data.....	23
Tabel 3. 4	Tabel <i>Dataset</i>	23
Tabel 3. 5	Hasil Matriks Jarak	24
Tabel 4. 1	<i>Atribut</i>	32
Tabel 4. 2	<i>Dataset</i> Pasien Penyakit <i>Leptospirosis</i>	32
Tabel 4. 3	Hasil Transformasi Data	33
Tabel 4. 4	Hasil Normalisasi Data	34
Tabel 4. 5	Data Pasien <i>Random</i>	34
Tabel 4. 6	Jarak <i>Euclidean</i> Setiap Pasangan Data	35
Tabel 4. 7	Jarak <i>Cluster</i> Terkecil.....	35
Tabel 4. 8	Jarak <i>Cluster</i> 5 dan 6 dengan <i>Cluster</i> yang Tersisa 1, 2, 3 dan 4	36
Tabel 4. 9	Matriks Jarak Pertama untuk <i>Single Linkage</i>	36
Tabel 4. 10	Jarak <i>Cluster</i> Terkecil Matriks Jarak Pertama.....	36
Tabel 4. 11	Jarak <i>Cluster</i> 1 dan 3 dengan <i>Cluster</i> yang Tersisa 2, 4 dan 56	36
Tabel 4. 12	Matriks Jarak Kedua untuk <i>Single Linkage</i>	37
Tabel 4. 13	Jarak <i>Cluster</i> Terkecil Matriks Jarak Kedua.....	37
Tabel 4. 14	Jarak <i>Cluster</i> 13, 56 dan 2 dengan <i>Cluster</i> yang Tersisa 4.....	37
Tabel 4. 15	Matriks Jarak Ketiga untuk <i>Single Linkage</i>	37
Tabel 4. 16	Jarak <i>Euclidean</i> Setiap Pasangan Data	38
Tabel 4. 17	Jarak <i>Cluster</i> Terkecil	38
Tabel 4. 18	Jarak <i>Cluster</i> 5 dan 6 dengan <i>Cluster</i> yang Tersisa 1, 2, 3 dan 4	39
Tabel 4. 19	Matriks Jarak Pertama untuk <i>Average Linkage</i>	39
Tabel 4. 20	Jarak <i>Cluster</i> Terkecil Matriks Jarak Pertama.....	39
Tabel 4. 21	Jarak <i>Cluster</i> 1 dan 3 dengan <i>Cluster</i> yang Tersisa 2, 4, dan 56	39
Tabel 4. 22	Matriks Jarak Kedua untuk <i>Average Linkage</i>	40
Tabel 4. 23	Jarak <i>Cluster</i> Terkecil Matriks Jarak Kedua.....	40

Tabel 4. 24	Jarak <i>Cluster</i> 1,3, 56 dan 2 dengan <i>Cluster</i> yang Tersisa 4.....	40
Tabel 4. 25	Matriks Jarak Ketiga untuk <i>Average Linkage</i>	40
Tabel 4. 26	Jarak <i>Euclidean</i> Setiap Pasangan Data	41
Tabel 4. 27	Jarak <i>Cluster</i> Terkecil	41
Tabel 4. 28	Jarak <i>Cluster</i> 5 dan 6 dengan <i>Cluster</i> yang Tersisa 1, 2, 3 dan 4	42
Tabel 4. 29	Matriks Jarak Pertama untuk <i>Complete Linkage</i>	42
Tabel 4. 30	Jarak <i>Cluster</i> Terkecil Matriks Jarak Pertama.....	42
Tabel 4. 31	Jarak <i>Cluster</i> 1 dan 3 dengan <i>Cluster</i> yang Tersisa 2, 4 dan 56	43
Tabel 4. 32	Matriks Jarak Kedua untuk <i>Complete Linkage</i>	43
Tabel 4. 33	Jarak <i>Cluster</i> Terkecil Matriks Jarak Kedua.....	43
Tabel 4. 34	Jarak <i>Cluster</i> 13, 56 dan 2 dengan <i>Cluster</i> yang Tersisa 4.....	43
Tabel 4. 35	Matriks Jarak Ketiga untuk <i>Complete Linkage</i>	43
Tabel 4. 36	<i>Cluster</i> 1 pada <i>Single Linkage</i> Skenario 2 <i>Cluster</i>	44
Tabel 4. 37	<i>Cluster</i> 2 pada <i>Single Linkage</i> Skenario 2 <i>Cluster</i>	45
Tabel 4. 38	<i>Cluster</i> 1 pada <i>Single Linkage</i> Skenario 3 <i>Cluster</i>	46
Tabel 4. 39	<i>Cluster</i> 2 pada <i>Single Linkage</i> Skenario 3 <i>Cluster</i>	46
Tabel 4. 40	<i>Cluster</i> 3 pada <i>Single Linkage</i> Skenario 3 <i>Cluster</i>	46
Tabel 4. 41	<i>Cluster</i> 1 pada <i>Single Linkage</i> Skenario 4 <i>Cluster</i>	47
Tabel 4. 42	<i>Cluster</i> 2 pada <i>Single Linkage</i> Skenario 4 <i>Cluster</i>	47
Tabel 4. 43	<i>Cluster</i> 3 pada <i>Single Linkage</i> Skenario 4 <i>Cluster</i>	48
Tabel 4. 44	<i>Cluster</i> 4 pada <i>Single Linkage</i> Skenario 4 <i>Cluster</i>	48
Tabel 4. 45	<i>Cluster</i> 1 pada <i>Single Linkage</i> Skenario 5 <i>Cluster</i>	49
Tabel 4. 46	<i>Cluster</i> 2 pada <i>Single Linkage</i> Skenario 5 <i>Cluster</i>	49
Tabel 4. 47	<i>Cluster</i> 3 pada <i>Single Linkage</i> Skenario 5 <i>Cluster</i>	49
Tabel 4. 48	<i>Cluster</i> 4 pada <i>Single Linkage</i> Skenario 5 <i>Cluster</i>	49
Tabel 4. 49	<i>Cluster</i> 5 pada <i>Single Linkage</i> Skenario 5 <i>Cluster</i>	50
Tabel 4. 50	<i>Cluster</i> 1 pada <i>Average Linkage</i> Skenario 2 <i>Cluster</i>	50
Tabel 4. 51	<i>Cluster</i> 2 pada <i>Average Linkage</i> Skenario 2 <i>Cluster</i>	51
Tabel 4. 52	<i>Cluster</i> 1 pada <i>Average Linkage</i> Skenario 3 <i>Cluster</i>	52
Tabel 4. 53	<i>Cluster</i> 2 pada <i>Average Linkage</i> Skenario 3 <i>Cluster</i>	52
Tabel 4. 54	<i>Cluster</i> 3 pada <i>Average Linkage</i> Skenario 3 <i>Cluster</i>	52
Tabel 4. 55	<i>Cluster</i> 1 pada <i>Average Linkage</i> Skenario 4 <i>Cluster</i>	53

Tabel 4. 56	<i>Cluster 2 pada Average Linkage Skenario 4 Cluster</i>	53
Tabel 4. 57	<i>Cluster 3 pada Average Linkage Skenario 4 Cluster</i>	54
Tabel 4. 58	<i>Cluster 4 pada Average Linkage Skenario 4 Cluster</i>	54
Tabel 4. 59	<i>Cluster 1 pada Average Linkage Skenario 5 Cluster</i>	55
Tabel 4. 60	<i>Cluster 2 pada Average Linkage Skenario 5 Cluster</i>	55
Tabel 4. 61	<i>Cluster 3 pada Average Linkage Skenario 5 Cluster</i>	55
Tabel 4. 62	<i>Cluster 4 pada Average Linkage Skenario 5 Cluster</i>	55
Tabel 4. 63	<i>Cluster 5 pada Average Linkage Skenario 5 Cluster</i>	56
Tabel 4. 64	<i>Cluster 1 pada Complete Linkage Skenario 2 Cluster</i>	56
Tabel 4. 65	<i>Cluster 2 pada Complete Linkage Skenario 2 Cluster</i>	57
Tabel 4. 66	<i>Cluster 1 pada Complete Linkage Skenario 3 Cluster</i>	58
Tabel 4. 67	<i>Cluster 2 pada Complete Linkage Skenario 3 Cluster</i>	58
Tabel 4. 68	<i>Cluster 3 pada Complete Linkage Skenario 3 Cluster</i>	58
Tabel 4. 69	<i>Cluster 1 pada Complete Linkage Skenario 4 Cluster</i>	59
Tabel 4. 70	<i>Cluster 2 pada Complete Linkage Skenario 4 Cluster</i>	60
Tabel 4. 71	<i>Cluster 3 pada Complete Linkage Skenario 4 Cluster</i>	60
Tabel 4. 72	<i>Cluster 4 pada Complete Linkage Skenario 4 Cluster</i>	60
Tabel 4. 73	<i>Cluster 1 pada Complete Linkage Skenario 5 Cluster</i>	61
Tabel 4. 74	<i>Cluster 2 pada Complete Linkage Skenario 5 Cluster</i>	61
Tabel 4. 75	<i>Cluster 3 pada Complete Linkage Skenario 5 Cluster</i>	62
Tabel 4. 76	<i>Cluster 4 pada Complete Linkage Skenario 5 Cluster</i>	62
Tabel 4. 77	<i>Cluster 5 pada Complete Linkage Skenario 5 Cluster</i>	62
Tabel 4. 78	<i>SSW Single Linkage 2 Cluster</i>	62
Tabel 4. 79	<i>SSB Single Linkage 2 Cluster</i>	63
Tabel 4. 80	<i>Rasio Single Linkage 2 Cluster</i>	63
Tabel 4. 81	<i>DBI Single Linkage 2 Cluster</i>	63
Tabel 4. 82	<i>SSW Single Linkage 3 Cluster</i>	63
Tabel 4. 83	<i>SSB Single Linkage 3 Cluster</i>	64
Tabel 4. 84	<i>Rasio Single Linkage 3 Cluster</i>	64
Tabel 4. 85	<i>DBI Single Linkage 3 Cluster</i>	64
Tabel 4. 86	<i>SSW Single Linkage 4 Cluster</i>	64
Tabel 4. 87	<i>SSB Single Linkage 4 Cluster</i>	65

Tabel 4. 88	<i>Rasio Single Linkage 4 Cluster</i>	65
Tabel 4. 89	<i>DBI Single Linkage 4 Cluster</i>	65
Tabel 4. 90	<i>SSW Single Linkage 5 Cluster</i>	65
Tabel 4. 91	<i>SSB Single Linkage 5 Cluster</i>	66
Tabel 4. 92	<i>Rasio Single Linkage 5 Cluster</i>	66
Tabel 4. 93	<i>DBI Single Linkage 5 Cluster</i>	66
Tabel 4. 94	<i>SSW Average Linkage 2 Cluster</i>	67
Tabel 4. 95	<i>SSB Average Linkage 2 Cluster</i>	67
Tabel 4. 96	<i>Rasio Average Linkage 2 Cluster</i>	67
Tabel 4. 97	<i>DBI Average Linkage 2 Cluster</i>	67
Tabel 4. 98	<i>SSW Average Linkage 3 Cluster</i>	67
Tabel 4. 99	<i>SSB Average Linkage 3 Cluster</i>	68
Tabel 4. 100	<i>Rasio Average Linkage 3 Cluster</i>	68
Tabel 4. 101	<i>DBI Average Linkage 3 Cluster</i>	68
Tabel 4. 102	<i>SSW Average Linkage 4 Cluster</i>	68
Tabel 4. 103	<i>SSB Average Linkage 4 Cluster</i>	69
Tabel 4. 104	<i>Rasio Average Linkage 4 Cluster</i>	69
Tabel 4. 105	<i>DBI Average Linkage 4 Cluster</i>	69
Tabel 4. 106	<i>SSW Average Linkage 5 Cluster</i>	70
Tabel 4. 107	<i>SSB Average Linkage 5 Cluster</i>	70
Tabel 4. 108	<i>Rasio Average Linkage 5 Cluster</i>	70
Tabel 4. 109	<i>DBI Average Linkage 5 Cluster</i>	70
Tabel 4. 110	<i>SSW Complete Linkage 2 Cluster</i>	71
Tabel 4. 111	<i>SSB Complete Linkage 2 Cluster</i>	71
Tabel 4. 112	<i>Rasio Complete Linkage 2 Cluster</i>	71
Tabel 4. 113	<i>DBI Complete Linkage 2 Cluster</i>	71
Tabel 4. 114	<i>SSW Complete Linkage 3 Cluster</i>	72
Tabel 4. 115	<i>SSB Complete Linkage 3 Cluster</i>	72
Tabel 4. 116	<i>Rasio Complete Linkage 3 Cluster</i>	72
Tabel 4. 117	<i>DBI Complete Linkage 3 Cluster</i>	72
Tabel 4. 118	<i>SSW Complete Linkage 4 Cluster</i>	73
Tabel 4. 119	<i>SSB Complete Linkage 4 Cluster</i>	73

Tabel 4. 120 Rasio <i>Complete Linkage 4 Cluster</i>	73
Tabel 4. 121 DBI <i>Complete Linkage 4 Cluster</i>	73
Tabel 4. 122 SSW <i>Complete Linkage 5 Cluster</i>	74
Tabel 4. 123 SSB <i>Complete Linkage 5 Cluster</i>	74
Tabel 4. 124 Rasio <i>Complete Linkage 5 Cluster</i>	74
Tabel 4. 125 DBI <i>Complete Linkage 5 Cluster</i>	75
Tabel 4. 126 Hasil Perbandingan Nilai DBI	75

DAFTAR GAMBAR

Gambar 2. 1 Hubungan Pengelompokan dengan <i>Hierarchical</i>	12
Gambar 2. 2 <i>Hierarchical Clustering</i>	14
Gambar 2. 3 Contoh Dendogram	14
Gambar 3. 1 Metode Penelitian	20
Gambar 3. 2 Dendogram <i>Single Linkage</i>	27
Gambar 3. 3 Dendogram <i>Average Linkage</i>	29
Gambar 3. 4 Dendogram <i>Complete Linkage</i>	31
Gambar 4. 1 Dendogram <i>Single Linkage</i> Skenario 2 Cluster	44
Gambar 4. 2 Dendogram <i>Single Linkage</i> Skenario 3 Cluster	45
Gambar 4. 3 Dendogram <i>Single Linkage</i> Skenario 4 Cluster	47
Gambar 4. 4 Dendogram <i>Single Linkage</i> Skenario 5 Cluster	48
Gambar 4. 5 Dendogram <i>Average Linkage</i> Skenario 2 Cluster.....	50
Gambar 4. 6 Dendogram <i>Average Linkage</i> Skenario 3 Cluster.....	51
Gambar 4. 7 Dendogram <i>Average Linkage</i> Skenario 4 Cluster.....	53
Gambar 4. 8 Dendogram <i>Average Linkage</i> Skenario 5 Cluster.....	54
Gambar 4. 9 Dendogram <i>Complete Linkage</i> Skenario 2 Cluster.....	56
Gambar 4. 10 Dendogram <i>Complete Linkage</i> Skenario 3 Cluster.....	58
Gambar 4. 11 Dendogram <i>Complete Linkage</i> Skenario 4 Cluster.....	59
Gambar 4. 12 Dendogram <i>Complete Linkage</i> Skenario 5 Cluster.....	61
Gambar 4. 13 Dendogram <i>Single Linkage</i> dari <i>Python</i>	80
Gambar 4. 14 Visualisasi <i>Agglomerative Hierarchical Clustering</i>	81

DAFTAR LAMPIRAN

Lampiran 1 <i>Dataset</i> Awal Pasien Penyakit <i>Leptospira</i>	89
Lampiran 2 <i>Dataset</i> Pasien Penyakit <i>Leptospira</i> setelah diseleksi.....	91
Lampiran 3 <i>Dataset</i> setelah ditransformasi	93
Lampiran 4 <i>Dataset</i> setelah dinormalisasi	95
Lampiran 5 <code>single_linkage.py</code>	97
Lampiran 6 <code>average_linkage.py</code>	100
Lampiran 7 <code>complete_linkage.py</code>	103

BAB I

PENDAHULUAN

Bab ini menyajikan latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup dan sistematika tugas akhir mengenai penerapan metode *Agglomerative Hierarchical Clustering* dalam pengelompokan data penyakit *leptospirosis*.

1.1. Latar Belakang

Leptospirosis adalah penyakit yang disebabkan oleh infeksi bakteri *patogen* yang disebut *Leptospira* dan ditularkan dari hewan kepada manusia (*zoonosis*) (Setiawan, 2008). Penularan bisa terjadi secara langsung akibat terjadi kontak langsung antara manusia (sebagai *host*) dengan urin atau jaringan binatang yang terinfeksi dan secara tidak langsung akibat terjadi kontak antara manusia dengan air, tanah atau tanaman yang terkontaminasi *urine* dari binatang yang terinfeksi *leptospira*. Jalan masuk yang biasa pada manusia adalah kulit yang terluka, terutama sekitar kaki, dan atau selaput mukosa di kelopak mata, hidung, dan selaput lendir (Ramadhani & Yuniyanto, 2012). *Leptospirosis* dapat menyebabkan peradangan hati akut, gejala penyakit ginjal, seperti penyakit kuning, pendarahan, dan *uremia*, dan gejala meningitis (Levett, 2001).

Penelitian mengenai penyakit *leptospirosis* di Indonesia secara terpusat dilakukan di Balai Besar Litbang Vektor dan Reservoir Penyakit (B2P2VRP) Salatiga, Jawa Tengah. Penelitian ini mengumpulkan sampel dari pasien yang memiliki gejala klinis *Leptospirosis* untuk diteliti ada tidaknya kandungan gen target *Leptospirosis* pada sampel tersebut. Penelitian ini menggunakan metode *Polymerase Chain Reaction* (PCR) dan *Quantitative Polymerase Chain Reaction* (qPCR) dalam mendeteksi *gen Leptospira*. Kedua metode tersebut menandai ada tidaknya *gen target Leptospira* di dalam sampel. Beberapa gen tersebut secara universal hadir dalam bakteri *Leptospira*, seperti gen 16S rRNA, atau RRS, SecY, lipL32, Lig A, dan lig B yang kemudian lebih efektif untuk diagnosis dini dalam spesimen klinis (Loureiro, et al., 2013). Eksperimen dengan menggunakan metode *Polymerase Chain Reaction* (PCR) dan *Quantitative Polymerase Chain Reaction* (qPCR) telah dilaksanakan oleh B2P2VRP Salatiga, akan tetapi ditemukan kesulitan dalam mengelompokkan data tersebut.

Data dapat dikelompokkan menggunakan mekanisme *Data Mining*. *Data Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari database yang besar (Pudjiantoro, et al., 2011). *Data Mining* adalah bagian dari *Knowledge Discovery in Database* (KDD). *Knowledge Discovery in Database* (KDD) adalah keseluruhan proses untuk mencari dan mengidentifikasi pola atau informasi data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti (Fayyad, 1996). *Clustering* merupakan salah satu metode dalam *data mining* yang bertujuan untuk mengelompokkan data yang mempunyai kemiripan dalam satu *cluster* (Agusta, 2007). *Clustering* adalah pengelompokan sejumlah data atau objek ke dalam *cluster* (*group*) sehingga setiap dalam *cluster* tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek dalam *cluster* yang lainnya (Alfina, et al., 2012). *Agglomerative Hierarchical Clustering* adalah metode *clustering* dalam penambangan data atau data mining (Kurniawan, 2017). *Agglomerative Hierarchical Clustering* adalah metode pengelompokan data yang dimulai dengan setiap satu pengamatan sebagai *clusternya* sendiri kemudian terus mengelompokkan pengamatan ke dalam kelompok yang semakin besar. Hal ini menghasilkan sebuah organisasi hirarki data yang dapat digunakan untuk mengelompokkan kumpulan data dan dari hirarki ini kelompok-kelompok tertentu dapat dipilih (Ariska, 2018).

Penelitian yang dilakukan oleh Soares, et al. untuk menganalisis distribusi spasial dan musiman *leptospirosis* dan mengidentifikasi kemungkinan ekologis dan sosial dari penularannya. Kejadian dan tingkat kematian berkolerasi dengan kondisi sosial ekonomi penduduk (Soares, et al., 2010). Penelitian lainnya dilakukan di Bantul, Yogyakarta oleh Sulistyawati, et al. Penelitian ini bertujuan untuk mengetahui distribusi kasus *leptospirosis* berdasarkan tingkat pendidikan dan jenis pekerjaan menggunakan pemetaan kasus, dan *clustering analysis*. Penelitian ini menemukan persentase penderita pria lebih besar dibandingkan wanita dan distribusi sesuai dengan jenis pekerjaan menemukan mayoritas penduduk yang bekerja sebagai petani memperoleh persentase tertinggi (Sulistyawati, et al., 2016). Penelitian juga dilakukan oleh Nery, et al dengan melakukan analisis secara klinis dan *epidemiologis* kasus *leptospirosis* untuk memprediksi penyakit menggunakan algoritma klasifikasi penambangan data. Penelitian ini menggunakan algoritma JRIP yang pendekatannya

berhasil memprediksi kasus *leptospirosis* (Nery, et al., 2017). Penelitian yang dilakukan oleh Panggabean untuk mengelompokan data penyakit *leptospirosis* menggunakan konsep *Data Mining*. Penelitian ini menggunakan *Principal Component Analysis* dan algoritma *K-Means* untuk pengelompokannya (Panggabean, 2019).

Berdasarkan uraian di atas, penulis tertarik untuk melakukan penelitian pengelompokan data penyakit *leptospirosis* dengan menggunakan metode *Agglomerative Hierarchical Clustering*.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka rumusan masalah yang dihadapi yaitu bagaimana menerapkan metode *Agglomerative Hierarchical Clustering* dalam pengelompokan data penyakit *leptospirosis*.

1.3. Tujuan dan Manfaat

Tujuan yang ingin dicapai dalam tugas akhir ini adalah dapat menerapkan metode *Agglomerative Hierarchical Clustering* dalam pengelompokan data penyakit *leptospirosis*.

Manfaat dilakukannya penelitian Tugas Akhir ini adalah untuk membantu mengelompokan data penyakit *leptospirosis*.

1.4. Ruang Lingkup

Ruang lingkup yang akan dibahas pada penelitian ini adalah:

1. Metode yang digunakan untuk mengelompokan data pasien penyakit *leptospirosis* adalah *Agglomerative Hierarchical Clustering*.
2. *Dataset* pasien penderita penyakit *leptospirosis* yang digunakan dalam penelitian ini diperoleh dari rekap data penelitian pasien di Balai Besar Litbang Vektor dan Reservoir Penyakit (B2P2VRP) Salatiga dengan tambahan dari RSUD Sunan Kalijaga Demak tahun 2019 berjumlah 72 data pasien.
3. Jumlah *cluster* yang digunakan pada penelitian ini adalah 4 jumlah *cluster* yang berbeda yaitu 2 *cluster*, 3 *cluster*, 4 *cluster*, dan 5 *cluster* dengan metode kedekatan *single linkage*, *average linkage*, dan *complete linkage*.
4. Mengukur kualitas hasil *clustering* yang didapatkan oleh metode *Agglomerative Hierarchical Clustering* dalam pengelompokan data penyakit *leptospirosis*.

1.5. Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penyusunan skripsi ini terdiri atas lima bab, yaitu pendahuluan, landasan teori, metodologi penelitian, hasil dan pembahasan, dan penutup.

BAB I PENDAHULUAN

Bab ini menyajikan latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup dan sistematika tugas akhir mengenai penerapan metode *Agglomerative Hierarchical Clustering* dalam pengelompokan data penyakit *leptospirosis*.

BAB II LANDASAN TEORI

Bab ini menyajikan tinjauan pustaka yang berhubungan dengan topik tugas akhir. Tinjauan pustaka yang digunakan dalam penyusunan tugas akhir ini meliputi penjelasan tentang *leptospirosis*, *clustering*, algoritma *Agglomerative Hierarchical Clustering*, *Davies Bouldin Index*, serta penelitian terkait.

BAB III METODOLOGI PENELITIAN

Bab ini menyajikan metode penelitian yang diacu terkait dalam tugas akhir penerapan metode *Agglomerative Hierarchical Clustering* dalam pengelompokan data penyakit *leptospirosis*.

BAB IV PEMBAHASAN DAN HASIL

Bab ini menguraikan pembahasan dan hasil dari penelitian yang telah dilakukan mengenai penerapan metode *Agglomerative Hierarchical Clustering* dalam pengelompokan data penyakit *leptospirosis*. Dalam bab ini dibahas mengenai proses-proses perhitungan sesuai dengan metodologi penelitian hingga hasil dari penelitian yang dilakukan.

BAB V PENUTUP

Bab ini berisi kesimpulan yang dapat diambil berkaitan dengan penelitian dan saran-saran untuk penelitian lebih lanjut.