

PERBANDINGAN METODE K-MEANS DAN SINGLE LINKAGE DALAM CLUSTER DATA

*Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika Fakultas Ilmu Komputer
Universitas Sriwijaya*



Oleh :

EZIL DESFA

NIM : 09021181320038

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2018**

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari selasa 23 Februari 2018 telah dilaksanakan sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

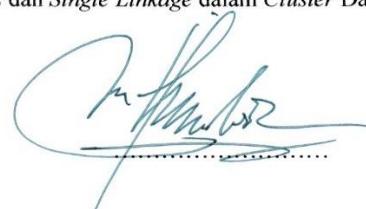
Nama : Ezil Desfa

NIM : 09021181320038

Judul : Perbandingan Metode *K-Means* dan *Single Linkage* dalam *Cluster Data*

1. Pembimbing I

Ir. M. Ihsan Jambak, M. Sc
NIP 196804052013081201



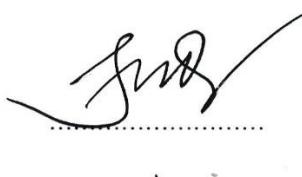
2. Pembimbing II

Danny Matthew Saputra, M.Sc
NIP 198505102015041002



3. Pengaji I

Rusdi Efendi, M.Kom
NIP 198201022011021201



4. Pengaji II

Hadipurnawan Satria, Ph.D
NIP 198004182015109101



Mengetahui,
Ketua Jurusan Teknik Informatika


Rifkie Primartha, M. T.
NIP 197706012009121004

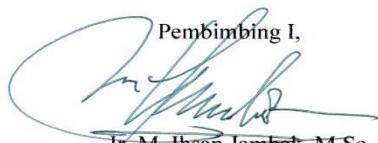
LEMBAR PENGESAHAN TUGAS AKHIR

PERBANDINGAN METODE K-MEANS DAN SINGLE LINKAGE DALAM
CLUSTER DATA

Oleh :

EZIL DESFA
NIM : 09021181320038

Indralaya, Maret 2018

Pembimbing I,

Ir. M. Ihsan Jambak, M.Sc
NIP 196804052013081201

Pembimbing II,

Danny Matthew Saputra, M.Sc
NIP 198505102015041002

Mengetahui,
Ketua Jurusan Teknik Informatika


Rifkie Primartha, M.T
NIP 197706012009121004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Ezil Desfa
NIM : 09021181320038
Program Studi : Teknik Informatika
Judul Skripsi : Perbandingan Metode *K-Means* dan
Single Linkage dalam *Cluster Data*
Hasil Pengecekan Software *iThenticate/Turnitin* : 2 %

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Palembang, Januari 2018



Ezil Desfa
NIM. 09021181320038

Motto

- ❖ Tidak penting seberapa lambat anda melaju, selagi anda tidak berhenti.
(Confucius)
- ❖ Meskipun kita tidak ada yang bisa kembali dan memulai yang baru. Tapi siapapun bisa memulai dari sekarang dan membuat akhir yang baru. (Carl Bard)

Kupersmbahkan Hasil Karyaku Kepada :

- ❖ Keluagaku
- ❖ Sahabat dan Teman-Teman IF Reguler

2013

Perbandingan Metode *K-Means* dan *Single Linkage* dalam *Cluster Data*

Oleh:

Ezil Desfa

09021181320038

ABSTRAK

K-Means dan *Single Linkage* merupakan metode yang digunakan untuk *clustering* atau pengelompokkan. Data yang digunakan berupa data bunga iris dengan jumlah *cluster* k adalah tiga. Metode *K-Means* dan *Single Linkage* memiliki perbedaan dimana metode *K-Means* dalam pengelompokkannya menggunakan titik *centroid*, sedangkan metode *Single Linkage* tidak menggunakan titik *centroid* tetapi menggunakan penggabungan jarak terdekat. Untuk melihat pengaruh titik *centroid* terhadap hasil pengelompokkan kedua metode tersebut maka dilakukan penelitian tentang perbandingan metode *K-Means* dan *Single Linkage* dengan mempertimbangkan performansi dan waktu komputasi. Berdasarkan hasil pengujian tersebut diketahui bahwa metode *K-Means* lebih baik dari metode *Single Linkage*.

Kata Kunci : *Clustering, K-Means, Single Linkage, Euclidean Distance.*

A Comparison of K-Means and Single Linkage Methods on Data Clustering

By :

Ezil Desfa

09021181320038

ABSTRACT

K-Means and Single Linkage are methods used for clustering (grouping). The data used is iris data with k cluster is three. K-Means and Single Linkage methods have differences where the K-Means method use centroid point on grouping, while Single Linkage does not use centroid point but using a grouping of closest distance. In order to find out the effect of centroid point on results of the two grouping methods, then proposing a research on comparing K-Means and Single Linkage method by considering performance and computation time. Based on the test results it was known that K-Means method is better than Single Linkage method.

Keyword: Clustering, K-Means, Single Linkage, Euclidean Distance

KATA PENGANTAR

Segala puji dan syukur penulis ucapkan kepada Allah SWT, karena berkat rahmat dan karunia-Nya penulis dapat menyelesaikan tugas akhir ini. Tugas akhir dengan judul **“Perbandingan Metode K-Means dan Single Linkage dalam Cluster Data”** ini disusun untuk memenuhi salah satu persyaratan dalam menyelesaikan pendidikan program Strata-I pada Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Pada kesempatan ini penulis ingin mengucapkan terima kasih kepada pihak-pihak yang telah membantu dan memberikan dukungan, bimbingan, kerjasama dan motivasi kepada penulis dalam menyelesaikan tugas akhir ini, yaitu kepada :

1. Ayah saya, Bapak Awaluddin; Ibu saya, Ibu Mitri Elida; Kakak saya, Wilfa Elza; Adik-adik saya, Esi Tria Putri dan Wifa Agusra; serta seluruh keluarga besar penulis yang telah mendoakan dan memberikan dukungan, semangat serta motivasi yang tidak dapat penulis hitung dan tuliskan satu persatu.
2. Terima kasih kepada Bapak Jaidan Jauhari, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Rifkie Primatha, M.T, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya
4. Ibu Anggina Primanita, M. IT, selaku Sekretaris dan Pembimbing Akademik di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

5. Bapak Ir. M. Ihsan Jambak, M.Sc, dan Bapak Danny Matthew Saputra, M.Sc, selaku pembimbing tugas akhir di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Bapak Rusdi Efendi, M.Kom dan Bapak Hadipurnawan Satria, Ph.D selaku penguji yang telah memberikan saran dan kritik yang membangun.
7. Seluruh Dosen Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya yang telah memberikan ilmu pengetahuan kepada penulis selama kegiatan akademik berlangsung.
8. Seluruh Karyawan dan Karyawati Fasilkom UNSRI yang telah membantu dalam urusan administrasi selama kegiatan akademik.
9. Sahabat Penulis Widya Damayanti dan Meita Jayani sebagai team sukses untuk segala hal menyangkut perkuliahan dari awal semester hingga semester akhir.
10. Seluruh sahabat seperjuangan di Teknik Informatika Reguler maupun Bilingual 2013 yang memberikan motivasi dan semangat untuk menyelesaikan tugas akhir ini.
11. Untuk semua pihak yang telah membantu menyelesaikan tugas akhir ini yang tidak dapat disebutkan satu persatu.

Penulis menyadari dalam penyusunan tugas akhir ini masih banyak terdapat kesalahan yang tidak disadari karena kekhilafan penulis semata, maka kritik dan saran yang membangun sangat diharapkan untuk membuat kesempurnaan di masa mendatang. Semoga tulisan ini dapat bermanfaat bagi semuanya.

Indralaya, Februari 2018

Penulis

DAFTAR ISI

Halaman

HALAMAN JUDUL	i
LEMBAR PENGESAHAN TUGAS AKHIR	ii
TANDA LULUS SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN.....	iv
MOTTO	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	xi
DAFTAR TABEL	xvii
DAFTAR GAMBAR.....	xix
DAFTAR LAMPIRAN	xxi

BAB I PENDAHULUAN

1.1 Latar Belakang Masalah.....	I-1
1.2 Rumusan Masalah	I-3
1.3 Tujuan Penelitian.....	I-4
1.4 Manfaat Penelitian.....	I-4
1.5 Batasan Masalah	I-4

BAB II TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu	II-1
2.2 Data Iris	II-3

2.3 <i>K-Means</i>	II-4
2.2.1 Komponen <i>K-Means</i>	II-4
2.2.2 Algoritma <i>K-Means</i>	II-5
2.4 <i>Single Linkage</i>	II-6

BAB III METODOLOGI PENELITIAN

3.1 Unit Penelitian	III-1
3.2 Metode Pengumpulan Data	III-1
3.3 Tahapan Penelitian	III-1
3.3.1 Menentukan Ruang Lingkup dan Unit Penelitian	III-2
3.3.2 Menemukan Dasar Teori yang Berkaitan dengan Penelitian	III-2
3.3.3 Menetapkan Kriteria Pengujian	III-3
a. Pengelompokan menggunakan metode <i>K-Means</i>	III-4
b. Pengelompokan menggunakan metode <i>Single Linkage</i> ...	III-5
3.3.4 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian	
.....	III-6
3.3.5 Melakukan Pengujian Penelitian	III-7
3.3.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan	
Penelitian	III-8
3.4 Metode Pengembangan Perangkat Lunak	III-9
3.4.1 Fase Insepsi	III-9
3.4.2 Fase Elaborasi	III-10
3.4.3 Fase Konstruksi	III-10
3.4.4 Fase Transisi.....	III-11
3.5 Penjadwalan Penelitian	III-11

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1 Pendahuluan	IV-1
4.2 Fase Insepsi	IV-1
4.2.1 Pemodelan Bisnis	IV-1
4.2.2 Kebutuhan	IV-2
4.2.3 Analisis	IV-3
4.2.4 Implementasi	IV-3
4.3 Fase Elaborasi	IV-3
4.3.1 Pemodelan Bisnis	IV-4
1. Diagram <i>Use Case</i>	IV-5
a. Tabel Definisi Aktor	IV-5
b. Tabel Definisi <i>Use Case</i>	IV-6
2. Aktivitas Diagram	IV-6
3. <i>Sequence Diagram</i>	IV-8
4.3.2 Kebutuhan	IV-12
a. Fitur Masukkan Data	IV-12
b. Fitur Melakukan <i>Clustering</i> dengan <i>K-Means</i>	IV-12
c. Fitur Melakukan <i>Clustering</i> dengan <i>Single Linkage</i>	IV-12
4.3.3 Analisis dan Desain.....	IV-13
a. Analisis Kebutuhan	IV-14
b. Analisis Data	IV-14
c. Analisis Metode <i>K-Means</i>	IV-15
d. Analisis Metode <i>Single Linkage</i>	IV-15
4.3.4 Implementasi	IV-19

4.4 Fase Konstruksi.....	IV-20
4.4.1 Pemodelan Bisnis	IV-20
4.4.2 Kebutuhan	IV-22
4.4.3 Analisis dan Desain	IV-22
4.4.4 Implementasi	IV-22
4.4.4.1 Implementasi Kelas	IV-23
4.4.4.2 Rancangan Antarmuka	IV-26
4.5 Fase Transisi	IV-28
4.5.1 Pemodelan Bisnis	IV-28
4.5.2 Kebutuhan	IV-29
4.5.3 Analisis dan Desain	IV-30
4.5.3.1 Rencana Pengujian <i>Use Case</i> Memasukkan Data ...	IV-30
4.5.3.2 Rencana Pengujian <i>Use Case</i> Melakukan <i>Clustering</i> dengan <i>K-Means</i>	IV-30
4.5.3.3 Rencana Pengujian <i>Use Case</i> Melakukan <i>Clustering</i> dengan <i>Single Linkage</i>	IV-31
4.5.4 Implementasi	IV-32
4.5.4.1 Pengujian <i>Use Case</i> Memasukkan Data	IV-34
4.5.4.2 Pengujian <i>Use Case</i> Melakukan <i>Clustering</i> dengan <i>K-Means</i>	IV-34
4.5.4.3 Pengujian <i>Use Case</i> Melakukan <i>Clustering</i> dengan <i>Single Linkage</i>	IV-34
BAB V HASIL DAN ANALISA PENELITIAN	
5.1 Pendahuluan	V-1

5.2 Data Hasil Percobaan Penelitian <i>K-Means</i>	V-1
5.2.1 Data Hasil Uji <i>Clustering K-Means</i>	V-2
5.2.2 Tabel <i>Cofussion Matrix K-Means</i>	V-4
5.2.3 Nilai <i>Precision, Recall, F-Measure</i> dan <i>Rand Index K-Means</i>	
.....	V-7
5.3 Data Hasil Percobaan Penelitian Perbandingan Metode <i>K-Means</i> dan <i>Single Linkage</i>	V-7
5.3.1 Data Hasil Uji Perbandingan <i>Clustering K-Means</i> dan <i>Single Linkage</i>	V-8
5.3.2 Tabel <i>Confussion Matrix</i> Perbandingan <i>K-Means</i> dan <i>Single Linkage</i>	V-8
5.3.3 Perbandingan Nilai <i>Precision, Recall, F-Measure</i> dan <i>Rand Index</i> Metode <i>K-Means</i> dan <i>Single Linkage</i>	V-9
5.4 Waktu Komputasi.....	V-10
5.5 Analisis Penelitian.....	V-10
BAB VI KESIMPULAN DAN SARAN	
6.1 Pendahuluan	VI-1
6.2 Kesimpulan	VI-1
6.3 Saran	VI-2
DAFTAR PUSTAKA	xxi

DAFTAR TABEL

	Halaman
Tabel III-1 Tabel Cluster Bunga Iris	III-3
Tabel III-2 Rancangan Tabel Hasil Penelitian.....	III-7
Tabel III-3 Rancangan <i>Confussion Matrix</i>	III-7
Tabel III-4 Rancangan Tabel Hasil Penelitian Perbandingan Metode	III-8
Tabel III-5 Tabel Penjadwalan Penelitian dalam Bentuk Work Breakdown Structure (WBS).....	III-12
Tabel IV-1 Definisi Aktor <i>Use Case</i>	IV-5
Tabel IV-2 Definisi <i>Use Case</i>	IV-6
Tabel IV-3 Kebutuhan Fungsional.....	IV-13
Tabel IV-4 Kebutuhan Non Fungsional	IV-13
Tabel IV-5 Skenario Masukkan Data.....	IV-16
Tabel IV-6 Skenario Melakukan <i>Clustering</i> dengan <i>K-Means</i>	IV-17
Tabel IV-7 Skenario Melakukan <i>Clustering</i> dengan <i>Single Linkage</i>	IV-18
Tabel IV-8 Implementasi Kelas	IV-23
Tabel IV-9 Rencana Pengujian <i>Use Case</i> Masukkan Data	IV-30
Tabel IV-10 Rencana Pengujian <i>Use Case</i> Melakukan <i>Clustering</i> dengan <i>K-Means</i>	IV-31
Tabel IV-11 Rencana Pengujian <i>Use Case</i> Melakukan <i>Clustering</i> dengan <i>Single Linkage</i>	IV-32
Tabel IV-12 Pengujian <i>Use Case</i> Masukkan Data.....	IV-35
Tabel IV-13 Pengujian <i>Use Case</i> Melakukan Clustering dengan <i>K-Means</i> .	IV-35
Tabel IV-14 Pengujian <i>Use Case</i> Melakukan Clustering dengan <i>Single</i>	

<i>Linkage</i>	IV-37
Tabel V-1 Rata-Rata Hasil <i>Clustering K-Means</i> Sebanyak 150 Data.....	V-3
Tabel V-2 <i>Confussion Matrix</i> Hasil <i>Clustering K-Means</i> dengan Nilai k=2 ..	V-4
Tabel V-3 <i>Confussion Matrix</i> Hasil <i>Clustering K-Means</i> dengan Nilai k=3 ..	V-5
Tabel V-4 <i>Confussion Matrix</i> Hasil <i>Clustering K-Means</i> dengan Nilai k=4 ..	V-6
Tabel V-5 <i>Confussion Matrix</i> Hasil <i>Clustering K-Means</i> dengan Nilai k=5 ..	V-6
Tabel V-6 Evaluasi Nilai Rata-rata Performa <i>Clustering K-Means</i>	V-7
Tabel V-7 Hasil Perbandingan <i>Clustering</i> Metode <i>Single Linkage</i>	V-8
Tabel V-8 <i>Confussion Matrix</i> Hasil Perbandingan <i>Clustering K-Means</i> dan <i>Single Linkage</i> dengan Nilai k=3	V-9
Tabel V-9 Evaluasi Nilai Perbandingan Performa <i>Clustering K-Means</i> dan <i>Single Linkage</i>	V-10
Tabel V-10 Tabel Waktu Komputasi Metode <i>K-Means</i> dan <i>Single Linkage</i> .	V-10

DAFTAR GAMBAR

	Halaman
Gambar II-1 Tiga Jenis Bunga Iris	II-3
Gambar II-1 Bagian-Bagian Bunga	II-3
Gambar III-1 Diagram Tahapan Penelitian	III-2
Gambar III-2 <i>Flowchart K-Means</i>	III-5
Gambar III-3 <i>Flowchart Single Linkage</i>	III-6
Gambar III-4 Penjadwalan untuk Tahap Menentukan Dasar Teori yang Berkaitan dengan Penelitian dan Menentukan Kriteria Pengujian	III-17
Gambar III-5 Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Insepsi.....	III-18
Gambar III-6 Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Elaborasi .	III-19
Gambar III-7 Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Konstruksi	III-20
Gambar III-8 Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Transisi	III-21
Gambar III-9 Penjadwalan untuk Tahap Melakukan Pengujian Penelitian, Analisa Hasil Pengujian Penelitian dan Membuat Kesimpulan	III-22
Gambar IV-1 <i>Use Case</i> Pengelompokkan Secara Manual	IV-2
Gambar IV-2 Diagram <i>Use Case</i>	IV-5
Gambar IV-3 Diagram Aktivitas <i>Use Case</i> Memasukkan Data	IV-7

Gambar IV-4	Diagram Aktivitas <i>Use Case</i> Melakukan <i>Clustering</i> dengan <i>K-Means</i>	IV-7
Gambar IV-5	Diagram Aktivitas <i>Use Case</i> Melakukan <i>Clustering</i> dengan <i>Single Linkage</i>	IV-7
Gambar IV-6	<i>Sequence Diagram</i> Memasukkan Data	IV-9
Gambar IV-7	<i>Sequence Diagram</i> Melakukan <i>Custering</i> dengan Metode <i>K-Means</i>	IV-10
Gambar IV-8	<i>Sequence Diagram</i> Melakukan <i>Custering</i> dengan Metode <i>Single Linkage</i>	IV-11
Gambar IV-9	Kelas Diagram Perangkat Lunak.....	IV-21
Gambar IV-10	Rancangan Antarmuka <i>FormClustering</i>	IV-27
Gambar IV-11	Rancangan Antarmuka Tabbed Pane Metode <i>Single</i> <i>Linkage</i>	IV-27
Gambar IV-12	<i>Deployment Diagram</i>	IV-29
Gambar IV-13	Antarmuka <i>FormClustering</i> Perangkat Lunak	IV-33
Gambar IV-14	Antarmuka Tabbed Pane Metode <i>Single Linkage</i>	IV-33
Gambar V-1	Nilai <i>Confusion Matrik</i> pada Metode <i>K-Means</i> dan <i>Single</i> <i>Linkage</i> pada k=3.....	V-11
Gambar V-2	Perbandingan Performa <i>Clustering</i> Metode <i>K-Means</i> dan <i>Single</i> <i>Linkage</i>	V-12
Gambar V-3	Perbandingan Waktu Komputasi <i>Clustering K-Means</i> dan <i>Single Linkage</i>	V-13

DAFTAR LAMPIRAN

Halaman

LAMPIRAN I Titik <i>Centroid</i> Merotode <i>K-Means</i>	A-1
LAMPIRAN II Data Hasil <i>Clustering K-Means</i> dengan 30 kali Pengujian..	B-1
LAMPIRAN III Kode Program Perangkat Lunak	C-1

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Salah satu teknik dalam data mining yang bisa digunakan dalam mengelompokkan data yaitu *clustering*. *Clustering* adalah salah satu metode yang terkenal dalam data mining dimana setiap objek data akan dikelompokkan kedalam satu kelompok berdasarkan kemiripannya, dan yang lainnya akan dikelompokan pada kelompok yang lain (Crowder 2009). *Cluster* adalah kumpulan obyek data yang mirip satu sama lain dalam kelompok yang sama dan berbeda dengan obyek-obyek di kelompok lain.

Menurut Tan (2013) ada dua metode untuk melakukan *clustering*, yaitu *Partitioning* dan *Hierarchical Clustering*. Metode partisi atau *partitioning* merupakan pengelompokan yang mengurai kumpulan data ke dalam serangkaian *cluster* yang terpisah. *Hierarchical Clustering* adalah jenis pengelompokan yang menghasilkan hierarki kluster, biasanya disajikan sebagai dendrogram (struktur pohon). Salah satu metode partisi yang sering digunakan adalah *K-Means* dan pada *Hierarchical Clustering* terdapat beberapa metode, salah satunya adalah metode *Single Linkage*.

K-Means merupakan metode pengelompokan yang sederhana dan dapat digunakan dengan mudah. Dalam pengelompokkannya metode *K-Means* melakukan pengelompokan berdasarkan pada jarak terdekat dengan *centroidnya* dengan data yang dihitung menggunakan jarak *Euclidean*, dimana pada awalnya metode *K-Means* memilih berapa titik centeroid. *K-Means* mempunyai

kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien (Arai and Barakbah 2007). Namun, *K-Means* mempunyai kelemahan pada penentuan *centroid* awal *cluster* (nilai k) yang diinisialisasi secara *random* sehingga pengelompokan data yang dihasilkan dapat berbeda-beda (Han, Pei et al. 2011). Jika inisialisasi nilai *random* kurang baik akan menyebabkan partisi yang berbeda serta hasil yang didapat menjadi kurang optimal. Hasil *cluster* yang terbentuk dari metode *K-Means* sangatlah bergantung pada inisiasi *centroid* awal *cluster* yang diberikan (Santosa 2007).

Metode *Single Linkage* mengelompokkan berdasarkan pada jarak terdekat antar data. Dimulai dengan dua objek yang dipisahkan dengan jarak terpendek, maka keduanya akan ditempatkan pada *cluster* atau kelompok pertama, begitu seterusnya. Namun metode ini juga memiliki kekurangan yaitu dimana interpretasi atau hierarki yang rumit, bergantung pada pemilihan teknik *intercluster similarity* yang lebih dikenal dengan istilah *linkage* serta komputasi yang mahal. Beberapa kelemahan dari linkage tersebut adalah sensitif terhadap adanya *outlier*, kesulitan menangani variasi bentuk dan ukuran, dan memisahkan *cluster* yang besar. Metode Pautan Tunggal (*Single Linkage*) akan mengelompokan dua objek yang mempunyai jarak terpendek dahulu.

Beberapa penelitian mengenai *clustering* sebelumnya telah dilakukan, diantaranya perbandingan Algoritma *K-Means* dan *Expectation Maximization* (EM) untuk Clusterisasi Nilai Mahasiswa Berdasarkan Asal Sekolah (Mardiani 2014), dilihat dari nilai-nilai koefisien silhouette, *K-Means* lebih unggul dari EM karena lebih banyak mendekati nilai 1.

Kemudian penelitian mengenai perbandingan Analisis Klaster *K-Means* dan *K-Median* Pada Data Indikator Kemiskinan (Ruswandi 2011), hasil penelitian menunjukkan bahwa pengelompokkan *K-Means* lebih baik dibanding *K-Median*. Penelitian lainnya adalah DSS untuk Menganalisis pH Kesuburan Tanah Menggunakan Metode *Single Linkage* (Kusuma, Hasanah et al. 2014). Hasil yang didapatkan bahwa waktu pemrosesan yang dibutuhkan metode *Single Linkage* untuk pengambilan keputusan lebih cepat dibandingkan metode *Complete Linkage* dan *Average Linkage*.

Berdasarkan uraian diatas, penelitian ini akan menganalisis perbandingan metode *K-Means* dan *Single Linkage* dalam *cluster* data untuk mengetahui metode mana yang menghasilkan *cluster* terbaik yang diharapkan dapat membentuk *cluster-cluster* data dan memperlihatkan pola persebaran data.

1.2 Rumusan Masalah

Perumusan masalah yang diangkat dari penelitian ini adalah bagaimana pengaruh *centroid* terhadap hasil *cluster* dari metode *K-Means* dan *Single Linkage*.

Untuk menyelesaikan permasalahan yang disebutkan diatas, maka dalam penelitian ini dibatasi dalam beberapa pertanyaan penelitian (*research question*) antara lain :

1. Bagaimana mekanisme metode *K-Means* dan *Single Linkage* untuk *cluster* data.
2. Bagaimana mengukur kinerja dari kedua metode tersebut dalam *cluster* data.

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengetahui Kelebihan dan kekurangan metode *K-Means* dan *Single Linkage*.
2. Membandingkan metode *K-Means* dan *Single Linkage* dari segi performansi dan waktu pemrosesan.

1.4 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut :

1. Memahami metode *K-Means* dan *Single Linkage* sebagai metode untuk melakukan *clustering*.
2. Mengetahui perbandingan metode *K-Means* dan *Single Linkage* sehingga mendapatkan metode yang lebih baik untuk melakukan *clustering*.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah :

1. Menggunakan jarak Euclidean dalam perhitungan kemiripan antar obyek pada metode *K-Means* dan *Single Linkage*.
2. Hasil *cluster* digunakan sebagai pembanding antara metode *K-Means* dan *Single Linkage*.
3. Data yang digunakan adalah data bunga iris sebanyak 150 data.

DAFTAR PUSTAKA

- Arai, K. and A. R. Barakbah (2007). "Hierarchical *K-Means*: an algorithm for centroids initialization for *K-Means*." Reports of the Faculty of Science and Engineering 36(1): 25-31.
- Crowder, M. (2009). "Cluster Randomised Trials by Richard J. Hayes, Lawrence H. Moulton." International Statistical Review 77(2): 305-305.
- Han, J., et al. (2011). Data mining: concepts and techniques, Elsevier.
- Kusuma, A. P., et al. (2014). "DSS untuk Menganalisis pH Kesuburan Tanah Menggunakan Metode Single Linkage." Jurnal EECCIS 8(1): 61-66.
- Mardiani, M. (2014). "Perbandingan Algoritma *K-Means* dan EM untuk Clusterisasi Nilai Mahasiswa Berdasarkan Asal Sekolah." Creative Information Technology Journal 1(4): 316-325.
- Ruswandi, B. (2011). "Analisis klaster *K-Means* dan *K-Median* pada data indikator kemiskinan: studi kasus data indikator kemiskinan kabupaten di indonesia tahun 2009."
- Santosa, B. (2007). "Data Mining (Teori dan Aplikasi)." Graha Ilmu, Yogyakarta.
- Tan, P. N., et al. (2013). "Data mining cluster analysis: basic concepts and algorithms." Introduction to data mining.
- Hartini, E. (2004). "Metode clustering hirarki." Pusat Pengembangan Teknologi Informasi dan Komputasi BATAN.
- Jain, A. K. and R. C. Dubes (1988). Algorithms for *clustering* data, Prentice-Hall, Inc.
- Kusuma, A. P., et al. (2014). "DSS untuk Menganalisis pH Kesuburan Tanah Menggunakan Metode Single Linkage." Jurnal EECCIS 8(1): 61-66.

MacQueen, J. (1967). "Some methods for classification and analysis of multivariate observations." Proceedings of the fifth Berkeley symposium on mathematical statistics and probability 1(14): 281-297.

Mardiani, M. (2014). "Perbandingan Algoritma K-Means dan EM untuk Clusterisasi Nilai Mahasiswa Berdasarkan Asal Sekolah." Creative Information Technology Journal 1(4): 316-325.

Ruswandi, B. (2011). "Analisis *cluster* k-means dan k-median pada data indikator kemiskinan: studi kasus data indikator kemiskinan kabupaten di indonesia tahun 2009."