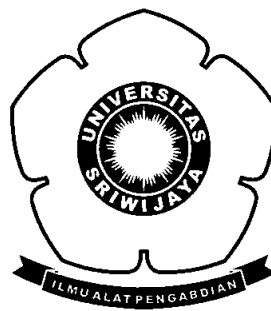


**PENGARUH PERHITUNGAN RENTANG DATA TERHADAP
PERFORMANSI ALGORITMA *CLUSTERING*
*SELF ORGANIZING MAP***

Diajukan sebagai syarat untuk menyelesaikan
Pendidikan Program Strata-1 pada
Jurusan Teknik Informatika



Oleh :

Tri Kurnia Sari
09021281520105

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2018**

LEMBAR PENGESAHAN TUGAS AKHIR

**PENGARUH PERHITUNGAN RENTANG DATA TERHADAP
PERFORMANSI ALGORITMA *CLUSTERING*
*SELF ORGANIZING MAP***

Oleh :

TRI KURNIA SARI
NIM : 09021281520105

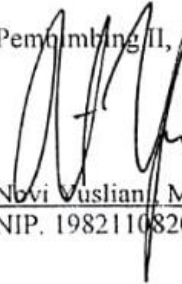
Indralaya, Januari 2019

Pembimbing I,



Dian Palupi Rini, M.Kom., Ph.D
NIP. 197802232006042002

Pembimbing II,



Novi Kusliani, M.T
NIP. 198211082012122001

Mengetahui,
Ketua Jurusan Teknik Informatika,



Rifkie Primartha, M.T
NIP. 197706012009121004

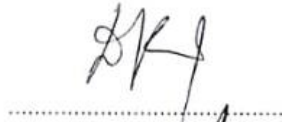
TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Rabu tanggal 12 Desember 2018 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Tri Kurnia Sari
NIM : 09021281520105
Judul : Pengaruh Perhitungan Rentang Data terhadap Performansi Algoritma
Clustering Self Organizing Map

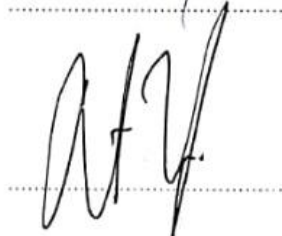
1. Pembimbing I

Dian Palupi Rini, M.Kom., Ph.D
NIP. 197802232006042002



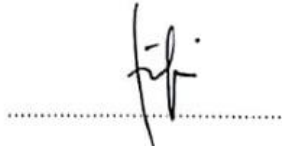
2. Pembimbing II

Novi Yusliani, M.T
NIP. 198211082012122001



3. Penguji I

Rifkie Primartha, M.T
NIP. 197706012009121004



4. Penguji II

Kanda Januar Miraswan, M.T
NIPUS. 1671080901900006



Mengetahui,
Ketua Jurusan Teknik Informatika


Rifkie Primartha, M.T
NIP. 197706012009121004



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Tri Kurnia Sari
NIM : 09021281520105
Program Studi : Teknik Informatika
Judul Skripsi : Pengaruh Perhitungan Rentang Data terhadap
Performansi Algoritma *Clustering Self Organizing Map*
Hasil Pengecekan Software *iThenticate/Turnitin* : 17 %

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 2019



Tri Kurnia Sari
NIM. 09021281520105

Motto :

- *Menggapai ridha Allah dan orang tua*
- *Sebaik-baik manusia adalah yang paling bermanfaat bagi manusia*
- *Beradab sebelum berilmu*

Kupersembahkan karya tulis ini kepada :

- *Orang tuaku tercinta dan saudariku tersayang*
- *Keluarga besarku*
- *Sahabat dan teman seperjuanganku*
- *Fakultas Ilmu Komputer*
- *Universitas Sriwijaya*

EFFECT OF DISTANCE MEASURES ON THE PERFORMANCE OF SELF ORGANIZING MAP CLUSTERING ALGORITHM

by :

Tri Kurnia Sari
09021281520105

ABSTRACT

Clustering is a valuable research field in data mining. Clustering groups data based on its proximity or similarity. An important component in clustering is the distance measurement between data points. The difference in distance measurement method of data will affect how close the distance between the data. Therefore, this research will examine the effect of distance measurement of data on the performance of clustering algorithm. This study uses Self Organizing Map algorithm. There are three distance measurement method are Euclidean distance, Manhattan distance, and Chebyshev distance. This research uses 4 data set are Iris, Wine, Car Evaluation, and Abalone. The cluster will be evaluated using Davies Bouldin Index. Self Organizing Map and Chebyshev distance produce better cluster than Euclidean distance and Manhattan distance. However, the computation time is longer than Euclidean distance and Manhattan distance.

Key Word : Clustering, Euclidean distance, Manhattan distance, Chebyshev distance, Self Organizing Map, Davies Bouldin Index

PENGARUH PERHITUNGAN RENTANG DATA TERHADAP PERFORMANSI ALGORITMA *CLUSTERING* *SELF ORGANIZING MAP*

Oleh :

Tri Kurnia Sari
09021281520105

ABSTRAK

Clustering merupakan bidang penelitian yang sangat penting dalam *data mining*. *Clustering* mengelompokkan data berdasarkan kedekatannya atau kemiripannya. Komponen yang sangat penting dalam *clustering* adalah pengukuran rentang antar data. Perbedaan metode perhitungan rentang data akan mempengaruhi seberapa dekatnya rentang antardata. Oleh sebab itu lah, penelitian ini akan meneliti tentang pengaruh perhitungan rentang data terhadap performansi algoritma *clustering*. Penelitian ini menggunakan algoritma *clustering Self Organizing Map*. Metode perhitungan rentang data yang digunakan ada 3, yaitu *Euclidean distance*, *Manhattan distance*, dan *Chebyshev distance*. *Data set* yang dijadikan objek penelitian ada 4, yaitu *Iris*, *Wine*, *Car Evaluation*, dan *Abalone*. Hasil *cluster* dievaluasi menggunakan *Davies Bouldin Index*. *Self Organizing Map* dan *Chebyshev distance* menghasilkan hasil *cluster* yang lebih baik dibandingkan *Euclidean distance* dan *Manhattan distance*. Namun, waktu komputasinya lebih lama dibandingkan kedua metode perhitungan rentang data tersebut.

Kata kunci : *Clustering*, *Euclidean distance*, *Manhattan distance*, *Chebyshev distance*, *Self Organizing Map*, *Davies Bouldin Index*

KATA PENGANTAR

Bismillahirrahmanir rahim

Puji syukur kepada Allah atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Penulis ingin menyampaikan rasa terima kasih kepada:

1. Orang tuaku, Ir. A. Rahman, M.S dan Dra. Tri Warsi Rahayu, saudariku, Mbak Rahmawati Rahayu, S.K.M dan Mbak Fitriyanti, S.T serta seluruh keluarga besarku yang selalu mendokan serta memberikan dukungan baik moril maupun materil.
2. Bapak Jaidan Jauhari, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya, Bapak Rifkie Primartha, M.T selaku Ketua Jurusan Teknik Informatika dan Ibu Anggina Primanita, M.IT selaku Sekretaris Jurusan Teknik Informatika.
3. Ibu Dian Palupi Rini, M.Kom., Ph.D selaku dosen pembimbing I dan Ibu Novi Yusliani, M.T selaku pembimbing II, yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
4. Bapak Osvari Arsalan, M.T selaku dosen pembimbing akademik, yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
5. Bapak Rifkie Primartha, M.T selaku dosen penguji I, dan Bapak Kanda Januar Miraswan, M.T selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
6. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Mbak Winda dan Kak Hafez serta seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
8. Achi, Pipit, Lifya, Ayu, Yulinda, Ipul serta seluruh teman-teman jurusan Teknik Informatika yang telah berbagi keluh kesah, motivasi, dan canda tawa di masa-masa perkuliahan ini.
9. BPH HMIF, Wifi, Digital Creative yang telah memberikan ruang bagi Penulis untuk berprestasi dan berkarya.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Indralaya, Januari 2019

Tri Kurnia Sari

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN LEMBAR PENGESAHAN	ii
HALAMAN TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah	I-1
1.3 Rumusan Masalah	I-5
1.4 Tujuan Penelitian	I-5
1.5 Manfaat Penelitian	I-6
1.6 Batasan Masalah	I-6
1.7 Sistematika Penulisan	I-7
1.8 Kesimpulan	I-8
BAB II KAJIAN LITERATUR	
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 Rentang Data	II-1
2.2.1.1 <i>Euclidean Distance</i>	II-2
2.2.1.2 <i>Manhattan Distance</i>	II-3
2.2.1.3 <i>Chebyshev Distance</i>	II-4
2.2.2 <i>Clustering</i>	II-5
2.2.2.1 <i>Self Organizing Map</i>	II-5
2.2.3 Evaluasi <i>Clustering</i>	II-10
2.2.3.1 <i>Davies Bouldin Index (DBI)</i>	II-10
2.2.4 Praproses Data	II-11
2.3 Penelitian Lain yang Relevan	II-13
2.4 Kesimpulan	II-26
BAB III METEDOLOGI PENELITIAN	
3.1 Pendahuluan	III-1
3.2 Unit Penelitian	III-1

3.3 Metode Pengumpulan Data	III-1
3.4 Tahapan Penelitian	III-2
3.4.1 Menetapkan Kerangka Kerjaa	III-3
3.4.2 Menetapkan Kriteria Pengujian	III-5
3.4.3 Menetapkan Format Data Pengujian	III-6
3.4.4 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian	III-9
3.4.5 Melakukan Pengujian Penelitian	III-9
3.4.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan Penelitian	III-10
3.5 Metode Pengembangan Perangkat Lunak	III-11
3.5.1 <i>Rational Unified Process</i> (RUP)	III-11
3.5.1.1 Fase Insepsi	III-12
3.5.1.2 Fase Elaborasi	III-13
3.5.1.3 Fase Konstruksi	III-13
3.5.1.4 Fase Transisi	III-14
3.6 Manajemen Proyek Penelitian	III-15
 BAB IV PENGEMBANGAN PERANGKAT LUNAK	
4.1 Pendahuluan	IV-1
4.2 <i>Rational Unified Process</i> (RUP)	IV-1
4.2.1 Fase Insepsi	IV-1
4.2.1.1 Pemodelan Bisnis	IV-1
4.2.1.2 Kebutuhan Sistem	IV-2
4.2.1.2.1 Fitur Praproses Data	IV-2
4.2.1.2.2 Fitur Proses Membangkitkan Nilai Bobot Acak	IV-3
4.2.1.2.3 Fitur <i>Clustering</i>	IV-3
4.2.1.3 Analisis dan Desain	IV-4
4.2.1.3.1 Analisis Kebutuhan Perangkat Lunak	IV-4
4.2.1.3.2 Analisis Data	IV-5
4.2.1.3.3 Analisis Praproses Data	IV-6
4.2.1.3.4 Analisis Perhitungan <i>Euclidean distance</i>	IV-8
4.2.1.3.5 Analisis Perhitungan <i>Manhattan distance</i>	IV-9
4.2.1.3.6 Analisis Perhitungan <i>Chebyshev distance</i>	IV-9
4.2.1.3.7 Analisis Algoritma <i>Self Organizing Map</i>	IV-10
4.2.1.3.8 Desain Perangkat Lunak	IV-11
4.2.2 Fase Elaborasi	IV-16
4.2.2.1 Pemodelan Bisnis	IV-17
4.2.2.1.1 Perancangan Data	IV-17
4.2.2.1.2 Perancangan Antarmuka	IV-17
4.2.2.2 Kebutuhan Sistem	IV-18
4.2.2.3 Diagram	IV-19
4.2.2.3.1 Diagram Aktivitas	IV-19
4.2.2.3.2 Diagram <i>Sequence</i>	IV-21
4.2.3 Fase Konstruksi	IV-24
4.2.3.1 Kebutuhan Sistem	IV-25
4.2.3.2 Diagram Kelas	IV-25

4.2.3.3 Kelas Analisis	IV-27
4.2.3.4 Implementasi	IV-28
4.2.3.4.1 Implementasi Kelas	IV-28
4.2.3.4.2 Implementasi Antarmuka	IV-29
4.2.4 Fase Transisi	IV-30
4.2.4.1 Pemodelan Bisnis	IV-30
4.2.4.2 Kebutuhan Sistem	IV-30
4.2.4.3 Rencana Pengujian	IV-31
4.2.4.3.1 Rencana Pengujian <i>Use Case</i> Melakukan Praproses Data	IV-31
4.2.4.3.2 Rencana Pengujian <i>Use Case</i> Melakukan Proses Membangkitkan Nilai Bobot Acak	IV-32
4.2.4.3.3 Rencana Pengujian <i>Use Case</i> Melakukan Proses <i>Clustering</i> Menggunakan Algoritma <i>Self Organizing Map</i> dan 3 Metode Perhitungan Rentang Data	IV-33
4.2.4.4 Implementasi	IV-33
4.2.4.4.1 Pengujian <i>Use Case</i> Melakukan Praproses Data	IV-34
4.2.4.4.2 Pengujian <i>Use Case</i> Melakukan Proses Membangkitkan Nilai Bobot Acak	IV-35
4.2.4.4.3 Pengujian <i>Use Case</i> Melakukan Proses <i>Clustering</i> Menggunakan Algoritma <i>Self Organizing Map</i> dan 3 Metode Perhitungan Rentang Data	IV-36
4.3 Kesimpulan	IV-38

BAB V HASIL DAN ANALISIS PENELITIAN

5.1 Pendahuluan	V-1
5.2 Data Hasil Percobaan/Penelitian	V-1
5.2.1 Konfigurasi Percobaan	V-1
5.2.2 Data Hasil Konfigurasi I	V-3
5.2.3 Data Hasil Konfigurasi II	V-4
5.2.4 Data Hasil Konfigurasi III	V-6
5.2.5 Data Hasil Konfigurasi IV	V-8
5.2.6 Data Hasil Konfigurasi V	V-9
5.3 Analisis Hasil Penelitian	V-13
5.4 Kesimpulan	V-39

BAB IV KESIMPULAN DAN SARAN

6.1 Pendahuluan	VI-1
6.1 Kesimpulan	VI-1
6.2 Saran	VI-3

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR TABEL

	Halaman
II-1 Hasil Percobaan dengan Menggunakan Algoritma <i>K-Means</i> , <i>Euclidean distance</i> , <i>Manhattan distance</i> , dan <i>Chebyshev distance</i>	II-14
II-2 Hasil Percobaan dengan Menggunakan Algoritma <i>K-Means</i> , <i>City Block distance</i> , <i>Euclidean distance</i> , <i>Cosine distance</i> dan <i>Correlation distance</i>	II-18
II-3 Hasil Percobaan dengan Menggunakan Algoritma <i>Single Linkage</i> , <i>Average Linkage</i> , dan <i>K-Means</i>	II-21
II-4 Hasil Percobaan dengan Menggunakan Algoritma <i>K-Medoids</i> dengan 7 Perhitungan Rentang Data	II-23
II-5 Perhitungan <i>Precision</i> untuk <i>Chebyshev distance</i> dan Minkowski <i>distance</i>	II-25
II-6 Perhitungan <i>Recall</i> untuk <i>Chebyshev distance</i> dan Minkowski <i>distance</i>	II-26
III-1 Rancangan Tabel Hasil <i>Clustering</i> Menggunakan Metode <i>Euclidean distance</i>	III-6
III-2 Rancangan Tabel Hasil <i>Clustering</i> Menggunakan Metode <i>Manhattan distance</i>	III-7
III-3 Rancangan Tabel Hasil <i>Clustering</i> Menggunakan Metode <i>Chebyshev distance</i>	III-7
III-4 Rancangan Tabel Hasil Evaluasi Internal pada Algoritma SOM dengan <i>Euclidean distance</i>	III-7
III-5 Rancangan Tabel Hasil Evaluasi Internal pada Algoritma SOM dengan <i>Manhattan distance</i>	III-8
III-6 Rancangan Tabel Hasil Evaluasi Internal pada Algoritma SOM dengan <i>Chebyshev distance</i>	III-8
III-7 Rancangan Tabel Hasil Perbandingan <i>Clustering</i>	III-10
IV-1 Kebutuhan Fungsional	IV-3
IV-2 Kebutuhan Non Fungsional	IV-4
IV-3 Contoh Data Uji Coba	IV-6
IV-4 Hasil Pembobotan Data Uji Coba	IV-6
IV-5 Hasil Mengganti Nilai yang Hilang	IV-7
IV-6 Hasil Normalisasi Data Uji	IV-8
IV-7 Contoh Data Bobot Awal	IV-8
IV-8 Hasil <i>Clustering</i> Algoritma <i>Self Organizing Map</i> dan <i>Euclidean distance</i>	IV-11
IV-9 Definisi Aktor	IV-12
IV-10 Definisi <i>Use Case</i>	IV-13
IV-11 Skenario Melakukan Praproses Data	IV-14
IV-12 Skenario Proses Membangkitkan Nilai Bobot Acak	IV-15
IV-13 Skenario Proses <i>Clustering</i> dengan Algoritma SOM dan 3 Metode Perhitungan Rentang Data	IV-16

IV-14 Implementasi Kelas	IV-28
IV-15 Rencana Pengujian <i>Use Case</i> Melakukan Praproses Data	IV-31
IV-16 Rencana Pengujian <i>Use Case</i> Melakukan Proses Membangkitkan Nilai Bobot Acak	IV-32
IV-17 Rencana Pengujian <i>Use Case</i> Melakukan Proses <i>Clustering</i>	IV-33
IV-18 Pengujian <i>Use Case</i> Melakukan Praproses Data	IV-34
IV-19 Pengujian <i>Use Case</i> Melakukan Proses Membangkitkan Nilai Bobot Acak	IV-35
IV-20 Pengujian <i>Use Case</i> Melakukan Proses <i>Clustering</i>	IV-36
V-1 Pengujian <i>Data Set Iris</i> dengan Jumlah <i>Cluster</i> = 3	V-3
V-2 Pengujian <i>Data Set Wine</i> dengan Jumlah <i>Cluster</i> = 3	V-4
V-3 Pengujian <i>Data Set Car Evaluation</i> dengan Jumlah <i>Cluster</i> = 3.....	V-6
V-4 Pengujian <i>Data Set Abalone</i> dengan Jumlah <i>Cluster</i> = 10	V-8
V-5 Perbandingan Nilai DBI dan Waktu Komputasi pada Iterasi ke-50....	V-9

DAFTAR GAMBAR

	Halaman
II-1 <i>Euclidean distance</i> di antara 2 titik	II-2
II-2 <i>Manhattan distance</i> di antara 2 titik	II-4
II-3 <i>Chebyshev distance</i> di antara 2 titik	II-5
II-4 Arsitektur Jaringan <i>Self Organizing Map</i> (SOM)	II-6
II-5 Hubungan Ketetanggaan pada Topologi <i>Linier Array</i>	II-8
II-6 Hubungan Ketetanggaan pada Topologi <i>Rectangular Grid</i>	II-9
II-7 Hubungan Ketetanggaan pada Topologi <i>Hexagonal Grid</i>	II-9
II-8 <i>Output K-Means</i> Menggunakan <i>Euclidean distance</i>	II-15
II-9 <i>Output K-Means</i> Menggunakan <i>Manhattan distance</i>	II-15
II-10 <i>Output K-Means</i> Menggunakan <i>Minkowski distance</i> dengan $P=4$	II-16
II-11 Grafik Perbandingan Distorsi pada <i>Euclidean K-Means</i> dan <i>Manhattan K-Means</i>	II-16
II-12 Grafik Perbandingan Distorsi pada <i>Minkowski K-Means</i> dengan Variasi Nilai P	II-16
II-13 Visualisasi <i>cluster</i> untuk <i>City Block distance</i>	II-18
II-14 Visualisasi <i>cluster</i> untuk <i>Euclidean distance</i>	II-18
II-15 Visualisasi <i>cluster</i> untuk <i>Cosine distance</i>	II-18
II-16 Visualisasi <i>cluster</i> untuk <i>Correlation distance</i>	II-18
III-1 Diagram Tahapan Penelitian	III-2
III-2 Diagram Tahapan Pengujian	III-9
III-3 Arsitektur <i>Rational Unified Process</i> (RUP)	III-11
III-4 Penjadwalan Penelitian Tahap Menentukan Ruang Lingkup dan Unit Penelitian	III-20
III-5 Penjadwalan Penelitian Tahap Menentukan Dasar Teori yang Berkait- an dengan Penelitian	III-21
III-6 Penjadwalan Penelitian Tahap Menentukan Kriteria Pengujian	III-21
III-7 Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Insepsi	III-22
III-8 Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Elaborasi	III-23
III-9 Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Konstruksi	III-24
III-10 Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Transisi	III-25
III-11 Penjadwalan Penelitian Tahap Melakukan Pengujian Penelitian	III-25
III-11 Penjadwalan Penelitian Tahap Membuat Analisa Hasil Pengujian dan Membuat Kesimpulan	III-25
IV-1 Contoh Perhitungan Mencari Nilai <i>Mean</i>	IV-7
IV-2 Contoh Perhitungan <i>Min-Max Normalization</i>	IV-7
IV-3 Contoh Perhitungan <i>Euclidean distance</i>	IV-8
IV-4 Contoh Perhitungan <i>Manhattan distance</i>	IV-9

IV-5	Contoh Perhitungan <i>Chebyshev distance</i>	IV-10
IV-6	Contoh Perhitungan Algoritma <i>Self Organizing Map</i> dan <i>Euclidean distance</i>	IV-11
IV-7	Diagram <i>Use Case</i>	IV-12
IV-8	Rancangan Antarmuka	IV-18
IV-9	Diagram Aktivitas Praproses Data	IV-19
IV-10	Diagram Aktivitas Proses Membangkitkan Nilai Bobot Acak	IV-20
IV-11	Diagram Aktivitas Proses <i>Clustering</i> dengan SOM dan 3 Metode Perhitungan Rentang Data	IV-21
IV-12	Diagram <i>Sequence</i> Praproses	IV-22
IV-13	Diagram <i>Sequence</i> Proses Pembangkitan Nilai Bobot Acak	IV-23
IV-14	Diagram <i>Sequence</i> Proses <i>Clustering</i>	IV-24
IV-15	Diagram Kelas	IV-26
IV-16	Kelas Analisis Melakukan Praproses Data	IV-27
IV-17	Kelas Analisis Melakukan Pembangkitan Nilai Bobot Acak	IV-27
IV-18	Kelas Analisis Melakukan Proses <i>Clustering</i>	IV-28
IV-19	Tampilan Antarmuka Halaman Utama Perangkat Lunak	IV-30
V-1	Perbandingan Nilai DBI pada <i>Data Set Iris</i> dengan Jumlah <i>Cluster = 3</i>	V-13
V-2	Perbandingan Waktu Komputasi pada <i>Data Set Iris</i> dengan Jumlah <i>Cluster = 3</i>	V-13
V-3	Perbandingan Nilai DBI pada <i>Data Set Iris</i> dengan Jumlah <i>Cluster = 4</i>	V-14
V-4	Perbandingan Waktu Komputasi pada <i>Data Set Iris</i> dengan Jumlah <i>Cluster = 4</i>	V-15
V-5	Perbandingan Nilai DBI pada <i>Data Set Iris</i> dengan Jumlah <i>Cluster = 5</i>	V-16
V-6	Perbandingan Waktu Komputasi pada <i>Data Set Iris</i> dengan Jumlah <i>Cluster = 5</i>	V-16
V-7	Perbandingan Nilai DBI pada <i>Data Set Iris</i> dengan Jumlah <i>Cluster = 6</i>	V-17
V-8	Perbandingan Waktu Komputasi pada <i>Data Set Iris</i> dengan Jumlah <i>Cluster = 6</i>	V-18
V-9	Perbandingan Nilai DBI pada <i>Data Set Wine</i> dengan Jumlah <i>Cluster = 3</i>	V-19
V-10	Perbandingan Waktu Komputasi pada <i>Data Set Wine</i> dengan Jumlah <i>Cluster = 3</i>	V-19
V-11	Perbandingan Nilai DBI pada <i>Data Set Wine</i> dengan Jumlah <i>Cluster = 4</i>	V-20
V-12	Perbandingan Waktu Komputasi pada <i>Data Set Wine</i> dengan Jumlah <i>Cluster = 4</i>	V-21
V-13	Perbandingan Nilai DBI pada <i>Data Set Wine</i> dengan Jumlah <i>Cluster = 5</i>	V-22
V-14	Perbandingan Waktu Komputasi pada <i>Data Set Wine</i> dengan Jumlah <i>Cluster = 5</i>	V-22
V-15	Perbandingan Nilai DBI pada <i>Data Set Wine</i> dengan Jumlah	

	<i>Cluster = 6</i>	V-23
V-16	Perbandingan Waktu Komputasi pada <i>Data Set Wine</i> dengan Jumlah <i>Cluster = 6</i>	V-24
V-17	Perbandingan Nilai DBI pada <i>Data Set Car Evaluation</i> dengan Jumlah <i>Cluster = 3</i>	V-25
V-18	Perbandingan Waktu Komputasi pada <i>Data Set Car Evaluation</i> dengan Jumlah <i>Cluster = 3</i>	V-25
V-19	Perbandingan Nilai DBI pada <i>Data Set Car Evaluation</i> dengan Jumlah <i>Cluster = 4</i>	V-26
V-20	Perbandingan Waktu Komputasi pada <i>Data Set Car Evaluation</i> dengan Jumlah <i>Cluster = 4</i>	V-27
V-21	Perbandingan Nilai DBI pada <i>Data Set Car Evaluation</i> dengan Jumlah <i>Cluster = 5</i>	V-28
V-22	Perbandingan Waktu Komputasi pada <i>Data Set Car Evaluation</i> dengan Jumlah <i>Cluster = 5</i>	V-28
V-23	Perbandingan Nilai DBI pada <i>Data Set Car Evaluation</i> dengan Jumlah <i>Cluster = 6</i>	V-29
V-24	Perbandingan Waktu Komputasi pada <i>Data Set Car Evaluation</i> dengan Jumlah <i>Cluster = 6</i>	V-30
V-25	Perbandingan Nilai DBI pada <i>Data Set Abalone</i> dengan Jumlah <i>Cluster = 10</i>	V-31
V-26	Perbandingan Waktu Komputasi pada <i>Data Set Iris</i> dengan Jumlah <i>Cluster = 10</i>	V-31
V-27	Perbandingan Nilai DBI pada <i>Data Set Abalone</i> dengan Jumlah <i>Cluster = 20</i>	V-32
V-28	Perbandingan Waktu Komputasi pada <i>Data Set Iris</i> dengan Jumlah <i>Cluster = 20</i>	V-33
V-29	Perbandingan Nilai DBI pada <i>Data Set Abalone</i> dengan Jumlah <i>Cluster = 25</i>	V-34
V-30	Perbandingan Waktu Komputasi pada <i>Data Set Iris</i> dengan Jumlah <i>Cluster = 25</i>	V-34
V-31	Perbandingan Nilai DBI pada <i>Data Set Abalone</i> dengan Jumlah <i>Cluster = 29</i>	V-35
V-32	Perbandingan Waktu Komputasi pada <i>Data Set Iris</i> dengan Jumlah <i>Cluster = 29</i>	V-36
V-33	Perbandingan Nilai DBI pada <i>Data Set Abalone</i> dengan Jumlah <i>Cluster = 35</i>	V-37
V-34	Perbandingan Waktu Komputasi pada <i>Data Set Iris</i> dengan Jumlah <i>Cluster = 35</i>	V-37

DAFTAR LAMPIRAN

	Halaman
1. Pengujian <i>Data Set Iris</i>	L-2
2. Pengujian <i>Data Set Wine</i>	L-51
3. Pengujian <i>Data Set Car Evaluation</i>	L-100
4. Pengujian <i>Data Set Abalone</i>	L-149
5. <i>Source Code</i> Program	L-211

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini akan membahas mengenai penjelasan umum keseluruhan penelitian, diantaranya adalah latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah atau ruang lingkup, sistematika penulisan dan kesimpulan.

1.2 Latar Belakang Masalah

Clustering adalah proses mengelompokkan data berdasarkan kemiripan atau ketidaksamaan di antara keduanya (Chahar, Kumar, & Chhabra, 2014). Dengan kata lain, *clustering* membagi data ke dalam kelompok (*cluster*) yang berbeda-beda. Tujuan utama *clustering* adalah membagi sekumpulan data menjadi himpunan bagian yang berbeda (*cluster*) sehingga data yang berada dalam subset tertentu memiliki sifat yang sama, sedangkan data dalam subset berbeda menunjukkan sifat yang berbeda dari data di subset lain (Bora & Gupta, 2014). *Clustering* telah diterapkan di berbagai bidang ilmu seperti *data mining*, pengenalan pola, kompresi data, dan lain-lain.

Komponen penting dalam *clustering* adalah pengukuran rentang data (Kouser & Sunita, 2013). Rentang data dapat mempengaruhi bentuk, volume, dan orientasi kelompok karena beberapa titik data mungkin saling berdekatan satu sama lain sesuai dengan ukuran rentang data tertentu, namun juga dapat saling berjauhan

berdasarkan ukuran rentang data lainnya (Chahar et al., 2014). Pengukuran rentang data yang berbeda akan menghasilkan rentang antardata yang berbeda pula. Oleh karena itu, perbedaan pengukuran rentang data mempengaruhi hasil dari suatu algoritma *clustering*.

Penelitian tentang perbandingan performansi suatu algoritma dengan menerapkan pengukuran rentang data yang berbeda telah dilakukan oleh beberapa peneliti. Penelitian terdahulu menggunakan 3 metode pengukuran rentang data yaitu *Euclidean distance*, *Manhattan distance*, dan *Chebyshev distance* serta algoritma yang digunakan adalah algoritma *clustering K-Means* (Kouser & Sunita, 2013). Hasil yang didapatkan adalah algoritma *K-Means* dengan menggunakan *Chebyshev distance* memberikan hasil akurasi yang paling besar dibandingkan dengan *Euclidean distance* dan *Manhattan distance*, sedangkan jumlah iterasi pada *Chebyshev distance* jauh lebih besar dibandingkan dengan *Euclidean distance* dan *Manhattan distance*.

Penelitian serupa juga dilakukan, metode pengukuran rentang yang digunakan yaitu *Euclidean distance*, *Manhattan distance*, dan *Minkowski distance*. Untuk mengetahui pengaruh dari ketiga metode tersebut, dilakukan penelitian dengan menggunakan algoritma *clustering K-Means* (Singh, Yadav, & Rana, 2013). Berdasarkan penelitian tersebut didapatkan hasil bahwa algoritma *clustering K-Means* yang menggunakan metode *Euclidean distance* memberikan hasil yang terbaik, sementara itu kombinasi algoritma *clustering K-Means* dengan metode *Manhattan distance* memberikan hasil yang paling buruk.

Penelitian terkait selanjutnya, metode pengukuran rentang data yang digunakan ada 4, yaitu *City Block (Manhattan) distance*, *Euclidean distance*, *Cosine distance*, dan *Correlation distance* (Bora & Gupta, 2014). Algoritma yang digunakan pada penelitian tersebut ialah algoritma *clustering K-Means*. Hasil yang didapatkan pada penelitian tersebut adalah *City Block (Manhattan) distance* menunjukkan kinerja yang lebih baik dalam hal waktu komputasi yang lebih sedikit, sementara itu *Cosine distance* membutuhkan waktu komputasi yang lebih lama dibandingkan pengukuran rentang data lainnya.

Penelitian lainnya juga dilakukan dengan menggunakan 6 metode pengukuran rentang data yang berbeda, yaitu *Euclidean distance*, *Standardized Euclidean distance*, *Manhattan distance*, *Mahalonobis distance*, *Cosine distance*, dan *Correlation distance*. Algoritma yang digunakan untuk mengetahui pengaruh keenam metode pengukuran rentang data tersebut yaitu *Single Linkage*, *Average Linkage*, dan *K-Means*. Hasil percobaan menunjukkan bahwa pengukuran rentang yang berbeda menghasilkan nilai yang lebih baik untuk *data set* yang berbeda bila digunakan pada algoritma yang berbeda (Chahar et al., 2014).

Penelitian terdahulu selanjutnya menggunakan 7 metode pengukuran rentang data yang berbeda, yaitu *Euclidean distance*, *Manhattan distance*, *LPNorm distance*, *ArcCosine distance*, *Canberra distance*, *Kulczynski distance*, dan *Lorentzian distance* serta menggunakan algoritma *clustering K-Medoids* (Anand, Padmanabham, & Govardhan, 2015). Hasil yang didapatkan dari penelitian ini adalah *Manhattan distance* menjadi metode pengukuran rentang yang paling baik untuk algoritma *clustering K-Medoids*.

Penelitian terkait lainnya juga dilakukan menggunakan metode pengukuran rentang data *Chebyshev distance* dan *Minkowski distance* pada teknik *Content-based Image Retrieval* (CBIR). Didapatkan hasil bahwa *Chebyshev distance* memberikan performansi yang lebih tinggi dibandingkan *Minkowski distance* (Kirdat & Patil, 2016).

Algoritma *clustering* yang digunakan pada penelitian terdahulu menggunakan metode *clustering* berbasis partisi dan hirarki. Algoritma *K-Means* dan *K-Medoids* (PAM) merupakan metode *clustering* berbasis partisi (Han, Kamber, & Pei, 2012). Algoritma lainnya yang juga digunakan untuk mengetahui pengaruh perhitungan rentang data yaitu algoritma *Single Linkage* dan *Average Linkage* yang merupakan metode *clustering* berbasis hirarki (Han et al., 2012). Oleh karena itu, penelitian mengenai pengaruh perhitungan rentang data pada metode *clustering* lainnya, seperti metode *clustering* berbasis *neural network* juga perlu dilakukan. *Self Organizing Map* (SOM) merupakan salah satu metode *clustering* berbasis *neural network* (Yang, Ouyang, & Shi, 2012).

Berdasarkan penelitian yang telah dilakukan sebelumnya didapatkan hasil bahwa perbedaan metode pengukuran rentang data akan memberikan hasil yang berbeda untuk algoritma yang berbeda pula. Oleh karena itu, penulis akan melakukan penelitian mengenai pengaruh perhitungan rentang data terhadap performansi algoritma *clustering Self Organizing Map* (SOM).

1.3 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah bagaimana pengaruh perhitungan rentang data *Euclidean distance*, *Manhattan distance*, dan *Chebyshev distance* terhadap performansi algoritma *clustering Self Organizing Map* (SOM).

Untuk menyelesaikan masalah di atas maka disusun menjadi empat pertanyaan penelitian (*Research Questions*):

1. Bagaimana mengukur rentang data menggunakan metode *Euclidean distance*, *Manhattan distance*, dan *Chebyshev distance* ?
2. Bagaimana penggunaan metode *Euclidean distance*, *Manhattan distance*, dan *Chebyshev distance* pada algoritma *clustering Self Organizing Map* (SOM) ?
3. Apakah perhitungan rentang data *Euclidean distance*, *Manhattan distance*, dan *Chebyshev distance* memiliki pengaruh pada algoritma *clustering Self Organizing Map* (SOM) ?
4. Bagaimana performansi algoritma *clustering Self Organizing Map* (SOM) dengan metode pengukuran rentang data yang berbeda-beda ?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut.

1. Mengetahui cara pengukuran rentang data menggunakan metode *Euclidean distance*, *Manhattan distance*, dan *Chebyshev distance*.
2. Mengetahui penggunaan metode *Euclidean distance*, *Manhattan distance*, dan *Chebyshev distance* pada algoritma *clustering Self Organizing Map* (SOM).

3. Mengetahui pengaruh perhitungan rentang data *Euclidean distance*, *Manhattan distance*, dan *Chebyshev distance* terhadap algoritma *clustering Self Organizing Map* (SOM).
4. Mengetahui performansi algoritma *clustering Self Organizing Map* (SOM) dengan metode pengukuran rentang data yang berbeda-beda.

1.5 Manfaat Penelitian

Adapun manfaat yang ingin didapatkan dari penulisan dan pelaksanaan penelitian ini adalah mengetahui metode pengukuran rentang data yang tepat untuk algoritma *clustering Self Organizing Map* (SOM).

1.6 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Algoritma *clustering* yang digunakan pada penelitian ini adalah *Self Organizing Map* (SOM).
2. Metode pengukuran rentang data yang digunakan adalah *Euclidean distance*, *Manhattan distance*, dan *Chebyshev distance*.
3. Pengukuran performansi algoritma *clustering Self Organizing Map* (SOM) menggunakan *Davies Bouldin Index* (DBI).
4. Data yang dijadikan objek penelitian merupakan *data set* standar yang diambil dari *UCI Machine Learning Repository*, yaitu *data set Iris*, *Wine*, *Car Evaluation*, dan *Abalone* yang memiliki jumlah atribut dan tipe data yang berbeda-beda.

1.7 Sistematika Penulisan

BAB I. PENDAHULUAN

Bab ini akan membahas latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah atau ruang lingkup, sistematika penulisan dan kesimpulan.

BAB II. KAJIAN LITERATUR

Bab ini akan menjelaskan dasar-dasar teori yang digunakan dalam penelitian. Pembahasan dalam bab ini meliputi hasil penelitian terdahulu yang relevan dengan penelitian ini, penjelasan mengenai metode pengukuran rentang data seperti *Euclidean distance*, *Manhattan distance*, dan *Chebyshev distance*, penjelasan mengenai algoritma *clustering Self Organizing Map* (SOM) serta penjelasan lainnya yang berkaitan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab ini akan membahas tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini akan membahas perancangan perangkat lunak yang akan dibangun pada penelitian ini.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini akan menampilkan hasil pengujian berdasarkan langkah-langkah yang telah direncanakan. Analisis diberikan sebagai basis dari kesimpulan yang diambil dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari semua uraian-uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan berguna dalam pengembangan perangkat lunak ini selanjutnya.

1.8 Kesimpulan

Perhitungan rentang data memainkan peran yang sangat penting dalam proses *clustering*. Oleh karena itu, pada penelitian ini akan membahas mengenai pengaruh perhitungan rentang data *Euclidean distance*, *Manhattan distance*, dan *Chebyshev distance* terhadap performansi algoritma *clustering Self Organizing Map* (SOM). Data yang digunakan pada penelitian ini adalah *data set Iris*, *Wine*, *Car Evaluation*, dan *Abalone*.

DAFTAR PUSTAKA

- À S. R., & Sonika, À. (2014). Effectiveness of Data Preprocessing for Data Mining. *International Journal of Current Engineering and Technology*, 4(5), 3480–3483.
- Alasadi, S. A., & Bhaya, W. S. (2017). Review of Data Processing Techniques in Data Mining. *Journal of Engineering and Applied Sciences*, 12(16), 4102–4107.
- Ali, S., & Smith, K. A. (2006). On learning algorithm selection for classification. *Applied Soft Computing Journal*, 6(2), 119–138. <https://doi.org/10.1016/j.asoc.2004.12.002>
- Anand, S., Padmanabham, P., & Govardhan, A. (2015). Effect of Distance measures on Partitional Clustering Algorithms using Transportation Data. *International Journal of Computer Science and Information Technologies*, 6(6), 5308–5312.
- Awodele, O., & Jegede, O. (2009). Neural Networks and its Application in Engineering. *Proceedings of Informing Science & IT Education Conference (InSITE) 2009*, 1–13. Retrieved from <http://proceedings.informingscience.org/InSITE2009/InSITE09p083-095Awodele542.pdf>
- Bora, D. J., & Gupta, A. K. (2014). Effect of Different Distance Measures on the Performance of K-Means Algorithm: An Experimental Study in Matlab. *International Journal of Computer Science and Information Technologies*, 5(2), 2501–2506. <https://doi.org/0975-9645>
- Chahar, V., Kumar, D., & Chhabra, J. K. (2014). Impact of Distance Measures on The Performance of Clustering Algorithm, 243(December), 0–7. <https://doi.org/10.1007/978-81-322-1665-0>
- Dahal, S. (2015). Effect of Different Distance Measures in Result of Cluster Analysis.
- Fausett, L.V (1993). *Fundamental of Neural Network: Architectures, Algorithm, And Application*. Prentice Hall, 1st edition. ISBN-13: 978-0133341867.
- Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, 17(2–3), 107–145. <https://doi.org/10.1023/A:1012801612483>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. San Francisco, CA, Itd: Morgan Kaufmann, 745. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>
- Kirdat, T., & Patil, V. V. (2016). Application of Chebyshev Distance and Minkowski Distance to CBIR Using Color Histogram. *International Journal of Innovative Research in Technology*, 2(9), 28–31.
- Kohonen, T. (2001). Self-Organizing Maps. In *Springer Series in Information Sciences* (Vol. 30, p. 501). <https://doi.org/10.1007/978-3-642-56927-2>
- Kouser, K., & Sunita. (2013). A comparative study of K Means Algorithm by Different Distance Measures. *International Journal of Innovative Research in Computer and Communication Engineering*, 1(9), 2443–2447.

- Krutchén, P. (2000). *The Rational Unified Process, An Introduction*, 2nd edition.
- Liu, P., El-darzi, E., Lei, L., Vasilakis, C., & Chountas, P. (2005). An Analysis of Missing Data Treatment Methods and Their Application to Health Care Dataset. *Health Care*, 583–590.
- Mythili, S., & Madhiya, E. (2014). An Analysis on Clustering Algorithms in Data Mining. *International Journal of Computer Science and Mobile Computing*, 3(1), 334–340.
- Nawrin, S., Rahman, M. R., & Akhter, S. (2017). Exploring K-Means with Internal Validity Indexes for Data Clustering in Traffic Management System, 8(3).
- Pachghare, V. K., Patole, V. a., & Kulkarni, D. P. (2010). Self Organizing Maps to Build Intrusion Detection System. *International Journal of Computer Applications*, 1(8), 1–4. <https://doi.org/10.5120/191-328>
- Pande, S. R., Sambare, M. S. S., & Thakre, V. M. (2012). Data Clustering Using Data Mining Techniques. *International Journal of Advanced Research in Computer and Communication Engineering*, 1(8), 494–499.
- Pujari, A. K., Rajesh, K., & Reddy, D. Surash. (2015). Clustering Technique in Data Mining - Survey. *IETE Journal of Research*, 47(August), 18–28. <https://doi.org/10.1080/03772063.2001.11416199>
- Saxena, A., & Singh, M. (2016). Using Categorical Attributes for Clustering. *International Journal of Scientific Engineering and Applied Science (IJSEAS)*, (2), 324–329.
- Singh, A., Yadav, A., & Rana, A. (2013). K-means with Three different Distance Metrics. *International Journal of Computer Applications*, 67(10), 13–17. <https://doi.org/10.5120/11430-6785>
- Tamiselvi, R., Sivasakthi, B., & Kavitha, R. (2015). An Efficient Preprocessing and Postprocessing Techniques in Data Mining. *INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS AN EFFICIENT PREPROCESSING AND*, 3(4), 80–85.
- Yang, L., Ouyang, Z., & Shi, Y. (2012). A modified clustering method based on self-organizing maps and its applications. *Procedia Computer Science*, 9, 1371–1379. <https://doi.org/10.1016/j.procs.2012.04.151>