

**PENGARUH PENENTUAN CENTROID AWAL MENGGUNAKAN
ALGORITMA *ANT COLONY OPTIMIZATION* TERHADAP
HASIL PENGKLASTERAN DATA BERDIMENSI TINGGI**

*Diajukan untuk Menyusun Skripsi
di Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI*



Imas Syaibun Nisa'
09021181520005

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2020**

LEMBAR PENGESAHAN TUGAS AKHIR

**PENGARUH PENENTUAN *CENTROID* AWAL MENGGUNAKAN
ALGORITMA *ANT COLONY OPTIMIZATION* TERHADAP HASIL
PENGKLASTERAN DATA BERDIMENSI TINGGI**

Oleh :

IMAS SYAIBUN NISA'

NIM : 09021181520005

Mengetahui,

Indralaya, 12 Agustus 2020

Pembimbing I,

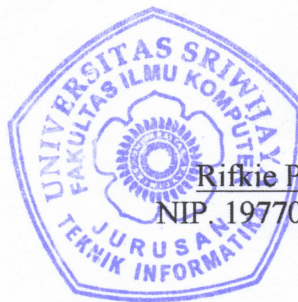


Ir. M. Ihsan Jambak, M.Sc., M.M

NIP. 196804052013081201

Mengetahui,

Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T

NIP. 197706012009121004


TANDA LULUS SIDANG TUGAS AKHIR

Pada hari Jumat, 24 Juli 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Imas Syaibun Nisa'
NIM : 09021181520005
Judul : Pengaruh Penentuan *Centroid* Awal Menggunakan Algoritma
Ant Colony Optimization Terhadap Hasil Pengklasteran Data
Berdimensi Tinggi

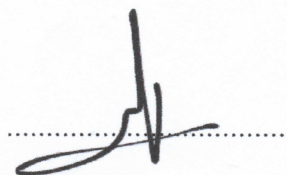
1. Pembimbing I

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



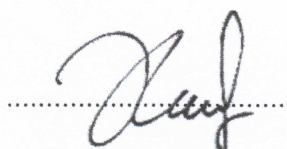
2. Penguji I

Dr. Abdiansah, S.Kom., M.CS.
NIP.198410012009121005



3. Penguji II

Kanda Januar Miraswan, M.T.
NIP. 199001092019031012



Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T
NIP. 197706012009121004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Imas Syaibun Nisa'

NIM : 09021181520005

Program Studi : Teknik Informatika

Judul Skripsi : Pengaruh Penentuan *Centroid* Awal Menggunakan Algoritma *Ant Colony Optimization* Terhadap Hasil Pengklasteran Data Berdimensi Tinggi

Hasil Pengecekan Software *iThenticate/Turnitin* : 5%

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.



Indralaya, 12 Agustus 2020



Imas Syaibun Nisa'

NIM. 09021181520005

HALAMAN MOTTO DAN PERSEMBAHAN

Motto :

Be happy, not because everything is good, but because you can see the good in everything.

Kupersembahkan karya tulis ini kepada :

- Kedua Orang tuaku tercinta
- Sahabat dan teman seperjuanganku
- Jurusan Teknik Informatika
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

THE EFFECT OF DETERMINE INITIAL CENTROID DETERMINATION
USING ANT COLONY OPTIMIZATION ALGORITHM ON HIGH
DIMENSIONAL DATA CLUSTERING RESULTS

Oleh:

Imas Syaibun Nisa'
09021181520005

ABSTRACT

Data Clustering can be done in several methods, one of which is the K-Means method. K-Means is a popular method of data clustering because of its simple implementation, it can handle large amounts of data and a relatively short process. In the k-means algorithm, initial centroid initialization is usually generated randomly, initial centroid initialization which is generated randomly often results in an optimum local solution. The optimum local solution has an impact on the quality of the clustering results being poor. Therefore it is necessary to find conditions that can meet the optimum global conditions where exploration occurs for the entire search space.

Ant Colony Optimization (ACO) is an ant algorithm in forming a colony. The ACO algorithm can avoid optimal local problems and have proven global solutions. In this research, an algorithm for ACO and K-Means-based data clustering is applied. The ACO algorithm can avoid optimal local problems and have proven global solutions. Based on the reliability of the ACO method in finding the optimum global solution, this paper tries to raise the ACO Algorithm as a reference for further action aimed at testing the effect of the ACO method in determining initial centroids on the quality of the results of the k-means clustering results

Keywords: *K-Means, Clustering, Ant Colony Optimization, Centroid, Optimum Local, Optimum Global.*

PENGARUH PENENTUAN CENTROID AWAL MENGGUNAKAN
ALGORITMA *ANT COLONY OPTIMIZATION* TERHADAP HASIL
PENGKLASTERAN DATA BERDIMENSI TINGGI

Oleh:

Imas Syaibun Nisa'
09021181520005

ABSTRAK

Klasterisasi Data (*Data Clustering*) dapat dilakukan dalam beberapa metode, salah satunya adalah metode *K-Means*. *K-Means* adalah salah satu metode klasterisasi data yang populer karena implementasi yang sederhana, dapat menangani data dalam jumlah besar dan prosesnya yang relatif singkat. Pada algoritma *k-means*, inisialisasi *centroid* awal biasanya dibangkitkan secara acak, inisialisasi *centroid* awal yang dibangkitkan secara acak sering kali menghasilkan solusi yang bersifat *local optimum*. Solusi yang bersifat *local optimum* berdampak pada kualitas hasil pengklasteran menjadi kurang baik. Maka dari itu perlu dicari kondisi yang dapat memenuhi kondisi *global optimum* dimana pengeplorasian terjadi untuk seluruh ruang pencarian

Ant Colony Optimization (ACO) merupakan suatu algoritma semut didalam membentuk suatu koloni. Algoritma ACO dapat menghindari dari permasalahan lokal optimal dan terbukti memiliki solusi global. Dalam penelitian ini diterapkan sebuah algoritma untuk klasterisasi data yang berbasis ACO dan *K-Means*. Algoritma ACO dapat menghindari dari permasalahan lokal optimal dan terbukti memiliki solusi global. Berdasarkan kehandalan metode ACO dalam menemukan solusi *global optimum* tersebut maka tulisan ini mencoba mengangkat Algoritma ACO sebagai acuan untuk dilakukan tindakan lanjutan yang bertujuan untuk menguji pengaruh metode ACO dalam penentuan *centroid* awal terhadap kualitas hasil pengklasteran algoritma *k-means*.

Kata Kunci: *K-means, Pengklasteran, Ant Colony Optimization, Pusat Klaster, Lokal Optimum, Global Optimum.*

KATA PENGANTAR



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

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

1. Kedua Orang tuaku tercinta, Robil Alamin dan Rohana, serta seluruh keluarga besarku yang selalu mendokan serta memberikan dukungan baik moril maupun materil.
2. Bapak Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Rifkie Primartha, S.T., M.T. selaku Ketua Jurusan Teknik Informatika.
4. Drs. Megah Mulya, M.T. selaku dosen pembimbing akademik yang telah membimbing, mengarahkan dan memberikan motivasi Penulis dalam proses perkuliahan.
5. Bapak Ir. M. Ihsan Jambak, M.Sc. selaku dosen pembimbing I yang telah membimbing, mengarahkan dan memberikan motivasi Penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.

6. Bapak Dr. Abdiansah, S.Kom., M.CS. selaku penguji I dan Bapak Kanda Januar Miraswan, M.T selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
7. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
8. Kak Ricy serta seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
9. Teman-teman seperjuanganku sampai akhir serta teman-teman seangkatan IF Reguler 2015 yang telah berjuang bersama dalam menimba ilmu di Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih banyak terdapat 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, 12 Agustus 2020

Imas Syaibun Nisa'

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN TUGAS AKHIR	iii
TANDA LULUS SIDANG TUGAS AKHIR	iv
HALAMAN PERNYATAAN	v
HALAMAN MOTTO DAN PERSEMBAHAN	vi
ABSTRACT	vii
ABSTRAKSI	viii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR TABEL	xv
DAFTAR GAMBAR	xix
DAFTAR PUSTAKA	xxi
BAB I PENDAHULUAN	
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian	I-5
1.5 Manfaat Penelitian	I-5
1.6 Batasan Masalah	I-6
1.7 Sistematika Penulisan	I-6
1.8 Kesimpulan	I-8
BAB II KAJIAN LITERATUR	
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1

2.2.1	<i>Clustering</i>	II-1
2.2.2	Algoritma <i>K-means</i>	II-2
2.2.3	Konversi Data	II-4
2.2.4	Reduksi Dimensi	II-6
2.2.5	Metode <i>Ant Colony Optimization</i>	II-9
2.2.6	Metode <i>Davies Bouldin Index</i>	II-17
2.3	Penelitian Lain yang Relevan	II-19
2.4	Kesimpulan	II-21
BAB III METODOLOGI PENELITIAN		
3.1	Pendahuluan	III-1
3.2	Unit Penelitian Data	III-1
3.3	Tahapan Penelitian	III-1
3.3.1	Menetapkan Kerangka Kerja	III-2
3.3.2	Menetapkan Kriteria Pengujian	III-7
3.3.3	Menetapkan Format Data Pengujian	III-9
3.3.4	Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian	III-12
3.3.5	Melakukan Pengujian Penelitian	III-12
3.3.6	Melakukan Pengujian Penelitian	III-13
3.3.6	Melakukan Analisa Hasil Pengujian dan Membuat kesimpulan Penelitian	III-13
3.4	Metode Pengembangan Perangkat Lunak	III-14
3.4.1	<i>Rational Unified Process (RUP)</i>	III-14
3.4.1.1	Fase Insepsi	III-15
3.4.1.2	Fase Elaborasi	III-15
3.4.1.3	Fase Konstruksi	III-16
3.4.1.4	Fase Transisi	III-16
3.5	Manajemen Proyek Penelitian	III-17
3.6	Kesimpulan	III-34
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 Sistem	IV-4
4.2.3 Analisis dan Desain	IV-15
4.3 Fase Elaborasi	IV-18
4.3.1 Pemodelan Bisnis	IV-18
4.3.2 Kebutuhan Sistem	IV-25
4.3.3 Analisis dan Desain	IV-31
4.4 Fase Konstruksi	IV-36
4.4.1 Pemodelan Bisnis	IV-37
4.4.2 Kebutuhan Sistem	IV-39
4.5 Fase Transisi	IV-40
4.5.1 Pemodelan Bisnis	IV-41
4.5.2 Kebutuhan Sistem	IV-41
4.5.3 Pengujian	IV-41
4.5.3.1 Rencana Pengujian	IV-42
4.6 Kesimpulan	IV-53
BAB V HASIL DAN ANALISIS PENELITIAN	V-1
5.1 Pendahuluan	V-1
5.2 Konfigurasi Percobaan	V-1
5.3 Hasil Pengujian	V-3
5.3.1 Hasil Pengujian k Optimum	V-3
5.3.2 Hasil Pengujian Pengklasteran <i>K-Means</i> dengan <i>Centroid</i> Awal Acak	V-4
5.3.3 Hasil Pengujian Parameter Algoritma ACO	V-6
5.3.3.1 Hasil Pengujian Jumlah Iterasi	V-6
5.3.3.2 Hasil Pengujian Jumlah Koloni Semut	V-8
5.3.3.3 Hasil Pengujian Nilai Intensitas (α)	V-10
5.3.4 Hasil Pengujian Pengklasteran <i>K-Means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma ACO	V-14
5.4 Olah Data	V-15

5.4.1 Olah Data Hasil Pengujian k Optimum	V-15
5.4.2 Olah Data Hasil Perbandingan Nilai DBI	V-16
5.4.3 Olah Data Hasil Perbandingan Waktu Komputasi	V-24
5.4.3 Olah Data Hasil Perbandingan Jumlah Iterasi	V-31
5.5 Analisis Hasil Pengujian	V-38
5.5.1 Analisis Hasil Pengujian Nilai k Optimum	V-38
5.5.2 Analisis Hasil Pengujian Parameter Algoritma ACO	V-38
5.5.3 Analisis Hasil Pengujian Pengklasteran k -means dengan Centroid Awal Acak pada Data Hasil Reduksi dan Tanpa Reduksi	V-39
5.5.4 Analisis Hasil Pengujian Pengklasteran k -means dengan Centroid Awal Acak dari Perhitungan Algoritma ACO pada Data Hasil Reduksi	V-40
5.5.5 Analisis Hasil Pengujian Pengklasteran k -means dengan Centroid Awal Acak pada Data Tanpa Reduksi	V-41
5.6 Kesimpulan	V-42

DAFTAR TABEL

	Halaman
III-1 Rancangan Penetapan Jumlah k Kluster Optimum	III-9
III-2 Rancangan Hasil Pengujian Jumlah Iterasi pada Data yang Belum dan Sudah Direduksi	III-10
III-3 Rancangan Hasil Pengujian dengan Jumlah Semut m pada Data yang Belum dan Sudah Direduksi	III-10
III-4 Rancangan Hasil Pengujian β dan α pada Data yang Belum dan Sudah Direduksi	III-11
III-5 Rancangan Hasil Pengujian <i>Clustering k-means</i> dengan <i>centroid</i> Awal secara Acak	III-11
III-6 Rancangan Hasil Pengujian <i>Clustering k-means</i> dengan <i>Centroid</i> Awal Menggunakan ACO Sebelum dan Sesudah Reduksi	III-12
III-7 Rancangan Hasil Perbandingan <i>k-means</i> dengan <i>Centroid</i> yang ditetapkan Menggunakan ACO dan <i>k-means</i> dengan <i>Centroid</i> yang Ditetapkan Acak Sebelum dan Sesudah Reduksi	III-14
III-8 Penjadwalan Penelitian dalam Bentuk <i>WBS</i>	III-18
IV-1 Definisi Aktor	IV-3
IV-2 Definisi <i>Use Case</i>	IV-3
IV-3 Kebutuhan Fungsional	IV-5
IV-4 Kebutuhan Non-Fungsional	IV-5
IV-5 Skenario <i>Use Case</i> Melakukan Konversi Data Teks ke Numerik	IV-6
IV-6 Skenario <i>Use Case</i> Melakukan Pengklasteran <i>K-means</i> dengan <i>Centroid</i> Awal secara Acak	IV-7
IV-7 Skenario <i>Use Case</i> Melakukan Pengklasteran <i>K-means</i> dengan <i>Centroid</i> Awal Algoritma ACO	IV-9
IV-8 Skenario <i>Use Case</i> Melakukan Pengklasteran <i>K-means</i> dengan <i>Centroid</i>	

Awal Acak Menggunakan Data Hasil Reduksi	IV-12
IV-9 Skenario <i>Use Case</i> Melakukan Pengklasteran <i>K-means</i> dengan <i>Centroid</i> Awal Algoritma ACO Menggunakan Data Hasil Reduksi	IV-14
IV-10 Implementasi Kelas	IV-39
IV-11 Rencana Pengujian <i>Use Case</i> Melakukan Konversi Data	IV-42
IV-12 Rencana Pengujian <i>Use Case</i> Melakukan Pengklasteran <i>K-means</i> dengan <i>Centroid</i> Awal Acak.....	IV-43
IV-13 Rencana Pengujian <i>Use Case</i> Melakukan Pengklasteran <i>K-means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma ACO.....	IV-43
IV-14 Rencana Pengujian <i>Use Case</i> Melakukan Pengklasteran <i>K-means</i> dengan <i>Centroid</i> Awal Acak Menggunakan Data Reduksi	IV-44
IV-15 Rencana Pengujian <i>Use Case</i> Melakukan Pengklasteran <i>K-means</i> dengan <i>Centroid</i> Awal dari Algoritma ACO Menggunakan Data Hasil Reduksi	IV-45
IV-16 Pengujian Fitur Konversi Data	IV-46
IV-17 Pengujian Fitur Pengklasteran <i>K-Means</i> dengan <i>Centroid</i> Awal Acak	IV-47
IV-18 Pengujian Fitur Pengklasteran <i>k-means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma ACO.....	IV-48
IV-19 Pengujian Fitur Pengklasteran <i>k-means</i> dengan <i>Centroid</i> Awal Acak Menggunakan Data Hasil Reduksi Dimensi	IV-50
IV-20 Pengujian Fitur Pengklasteran <i>K-means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma ACO Menggunakan Data Hasil Reduksi Dimensi	IV-51
V-1 Hasil Pengujian <i>k</i> Optimum	V-4
V-2 Hasil Pengujian Pengklasteran <i>-Means</i> dengan <i>Centroid</i> Awal Acak	V-5
V-3 Hasil Pengujian Jumlah Iterasi	V-7
V-4 Hasil Pengujian Jumlah Koloni Semut	V-9
V-5 Hasil Pengujian Nilai Intensitas (α).....	V-11
V-6 Hasil Pengujian Pengklasteran <i>K-Means</i> dengan <i>Centroid</i> Awal dari	

	Perhitungan Algoritma ACO.....	V-15
V-7	Deskripsi Nilai DBI	V-17
V-8	Hasil Uji Normalitas Nilai DBI	V-18
V-9	Hasil Uji Kruskal-Wallis Perbandingan Nilai DBI	V-18
V-10	Hasil Uji <i>Mann-Whitney</i> Perbandingan Nilai DBI pada Data Hasil Pengklastran <i>K-means</i> dan <i>K-means</i> +SVD	V-19
V-11	Hasil Uji <i>Mann-Whitney</i> Perbandingan Nilai DBI pada Data Hasil Pengklastran <i>K-means</i> dan <i>K-means</i> +ACO	V-20
V-12	Hasil Uji <i>Mann-Whitney</i> Perbandingan Nilai DBI pada Data Hasil Pengklastran <i>K-means</i> dan <i>K-means</i> +ACO+SVD	V-21
V-13	Hasil Uji <i>Mann-Whitney</i> Perbandingan Nilai DBI pada Data Hasil Pengklastran <i>K-means</i> +SVD dan <i>K-means</i> +ACO	V-21
V-14	Hasil Uji <i>Mann-Whitney</i> Perbandingan Nilai DBI pada Data Hasil Pengklastran <i>K-means</i> +SVD dan <i>K-means</i> +ACO+SVD	V-22
V-15	Hasil Uji <i>Mann-Whitney</i> Perbandingan Nilai DBI pada Data Hasil Pengklastran <i>K-means</i> +ACO dan <i>K-means</i> +ACO+SVD	V-21
V-16	Deskripsi Waktu Komputasi	V-24
V-17	Hasil Uji Normalitas Waktu Komputasi	V-25
V-18	Hasil Uji Kruskal-Wallis Perbandingan Waktu Komputasi	V-25
V-19	Hasil Uji <i>Mann-Whitney</i> Perbandingan Waktu Komputasi pada Data Hasil Pengklastran <i>K-means</i> dan <i>K-means</i> +SVD	V-26
V-20	Hasil Uji <i>Mann-Whitney</i> Perbandingan Waktu Komputasi pada Data Hasil Pengklastran <i>K-means</i> dan <i>K-means</i> +ACO	V-27
V-21	Hasil Uji <i>Mann-Whitney</i> Perbandingan Waktu Komputasi pada Data Hasil Pengklastran <i>K-means</i> dan <i>K-means</i> +ACO+SVD.....	V-28
V-22	Hasil Uji <i>Mann-Whitney</i> Perbandingan Waktu Komputasi pada Data Hasil Pengklastran <i>K-means</i> +SVD dan <i>K-means</i> +ACO	V-28
V-23	Hasil Uji <i>Mann-Whitney</i> Perbandingan Waktu Komputasi pada Data Hasil Pengklastran <i>K-means</i> +SVD dan <i>K-means</i> +ACO+SVD .	V-29
V-24	Hasil Uji <i>Mann-Whitney</i> Perbandingan Waktu Komputasi pada Data Hasil Pengklastran <i>K-means</i> +ACO dan <i>K-means</i> +ACO+SVD..	V-29

V-25	Deskripsi Jumlah Iterasi	V-31
V-26	Hasil Uji Normalitas Jumlah Iterasi	V-32
V-27	Hasil Uji Kruskal-Wallis Perbandingan Jumlah Iterasi	V-32
V-28	Hasil Uji <i>Mann-Whitney</i> Perbandingan Jumlah Iterasi pada Data Hasil Pengklasteran <i>K-means</i> dan <i>K-means+SVD</i>	V-33
V-29	Hasil Uji <i>Mann-Whitney</i> Perbandingan Jumlah Iterasi pada Data Hasil Pengklasteran <i>K-means</i> dan <i>K-means+ACO</i>	V-34
V-30	Hasil Uji <i>Mann-Whitney</i> Perbandingan Jumlah Iterasi pada Data Hasil Pengklasteran <i>K-means</i> dan <i>K-means+ACO+SVD</i>	V-35
V-31	Hasil Uji <i>Mann-Whitney</i> Perbandingan Jumlah Iterasi pada Data Hasil Pengklasteran <i>K-means+SVD</i> dan <i>K-means+ACO</i>	V-35
V-32	Hasil Uji <i>Mann-Whitney</i> Perbandingan Jumlah Iterasi pada Data Hasil Pengklasteran <i>K-means+SVD</i> dan <i>K-means+ACO+SVD</i> .	V-36
V-33	Hasil Uji <i>Mann-Whitney</i> Perbandingan Jumlah Iterasi pada Data Hasil Pengklasteran <i>K-means+ACO</i> dan <i>K-means+ACO+SVD</i> ..	V-37

DAFTAR GAMBAR

	Halaman
II-1 <i>Flowchart</i> Algoritma k-means	II-4
II-2 Matriks Dekomposisi <i>Singular Value Decomposition</i>	II-7
II-3 <i>Flowchart</i> Algoritma <i>Ant Colony Optimization</i>	II-16
II-4 Arsitektur <i>Rational Unified Process</i> (RUP)	II-18
III-1 Tahapan Penelitian	III-2
III-2 Kerangka Kerja Penelitian	III-3
III-3 Pengujian Penelitian	III-13
III-4 Penjadwalan Penelitian Tahap Menentukan Ruang Lingkup dan Unit Penelitian	III-22
III-5 Penjadwalan Penelitian Tahap Menentukan Dasar Teori yang Berkaitan dengan Penelitian	III-23
III-6 Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Insepsi	III-23
III-7 Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Elaborasi	III-24
III-8 Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Konstruksi	III-25
III-9 Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Transisi	III-26
III-10 Penjadwalan Penelitian Tahap Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan	III-27
IV-1 Diagram <i>Use Case</i>	IV-2
IV-2 Diagram <i>Sequence</i> Melakukan Konversi Data	IV-20
IV-3 Diagram <i>Sequence</i> Melakukan Pengklasteran <i>k-means</i> dengan <i>Centroid</i> Awal Acak	IV-21
IV-4 Diagram <i>Sequence</i> Melakukan Pengklasteran <i>k-means</i> dengan <i>Centroid</i>	

Awal dari Perhitungan Algoritma ACO	IV-22
IV-5 Diagram <i>Sequence</i> Melakukan Pengklasteran <i>k-means</i> dengan <i>Centroid</i> Acak Menggunakan Data yang Telah Direduksi	IV-23
IV-6 Diagram <i>Sequence</i> Melakukan Pengklasteran <i>k-means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma ACO Menggunakan Data yang Telah Direduksi	IV-24
IV-7 Rancangan Antar Muka Perangkat Lunak.....	IV-25
IV-8 Rancangan Fitur Buka Direktori.....	IV-25
IV-9 Rancangan Fitur Input Jumlah <i>Cluster</i>	IV-26
IV-10 Rancangan Input Parameter Metode ACO	IV-26
IV-11 Rancangan Fitur Metode <i>Clustering</i>	IV-27
IV-12 Rancangan Fitur Tombol Mulai Proses	IV-29
IV-13 Rancangan Fitur Menampilkan Waktu Komputasi, Jumlah Iterasi dan nilai DBI.....	IV-29
IV-14 Rancangan Fitur Menampilkan Pusat <i>cluster</i> akhir.....	IV-30
IV-15 Rancangan Fitur Menampilkan Hasil <i>Clustering</i> Dokumen	IV-30
IV-16 Diagram Aktivitas Konversi Data	IV-32
IV-17 Diagram Aktivitas Melakukan Pengklasteran <i>K-means</i> dengan <i>Centroid</i> Awal Acak	IV-33
IV-18 Diagram Aktivitas Melakukan Pengklasteran <i>K-means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma ACO	IV-34
IV-19 Diagram Aktivitas Melakukan Pengklasteran <i>K-means</i> dengan <i>Centroid</i> Awal Acak Menggunakan Data Hasil Reduksi	IV-35
IV-20 Diagram Aktivitas Melakukan Pengklasteran <i>K-means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma ACO Menggunakan Data Hasil Reduksi	IV-38
IV-21 Diagram Kelas	IV-38
V-1 Kurva Hasil Penentuan <i>k</i> Optimum	V-1

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini membahas latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penulisan dan kesimpulan. Bab ini akan memberikan penjelasan umum mengenai keseluruhan penelitian. Pendahuluan dimulai dengan penjelasan mengenai proses pengklasteran data menggunakan algoritma *k-Means* yang kemudian dilanjutkan dengan penelitian yang berkaitan dengan penggunaan algoritma *Ant Colony Optimization* dalam penentuan titik pusat awal klaster yang akan digunakan pada pengklasteran Dokumen Berdimensi Tinggi.

1.2 Latar Belakang Masalah

Pengklasteran adalah salah satu metode mayor pada *data mining* yang mengidentifikasi secara eksploratif dan menarik pola dari kumpulan data yang besar (Sunil dan Balachandran, 2013). Secara konsep pengklasteran adalah proses pengelompokan suatu himpunan objek data kedalam k klaster, sehingga objek-objek dalam satu kelompok memiliki kesamaan yang tinggi (*similarity*) dan memiliki perbedaan yang tinggi dengan objek-objek pada kelompok lain (*dissimilarity*) (Han, Kamber, dan Pei, 2012). Pengklasteran dikenal sebagai pembelajaran tanpa pengawasan (*unsupervised learning*) karena informasi label pada objek tidak ada. Menurut Santoso (2007), ada dua pendekatan dalam pengklasteran yaitu partisi dan hirarki. Pengklasteran telah menjadi masalah yang

secara luas dipelajari dalam berbagai disiplin ilmu termasuk diantaranya jaringan saraf tiruan, kecedasan buatan, dan statistik (Saranya dan Krishnakumari, 2011).

Algoritma *k-means* adalah salah satu algoritma pengklasteran dengan pendekatan partisi yang dikembangkan oleh *Mac Queen* pada tahun 1967. *K-means* populer digunakan dalam pengklasteran karena prosesnya yang sederhana, cepat dan efisien dalam komputasi dan kompleksitasnya linier dengan jumlah data (Santoso, 2007; Saranya dan Krishnakumari, 2011; Mishra, Saini, dan Bagri, 2015).

Namun, kualitas hasil pengklasteran algoritma *k-means* sangat tergantung pada inisialisasi *centroid* awal (Ahmed dan Rahim, 2017). Pada algoritma *k-means*, inisialisasi *centroid* awal biasanya dibangkitkan secara acak, inisialisasi *centroid* awal yang dibangkitkan secara acak sering kali menghasilkan solusi yang bersifat *local optimum*, yaitu dimana kondisi *optimum* hanya terjadi pada lingkup kluster atau *optimum* pada *similarity* dalam konsep pengklasteran kondisi ini terjadi karena pengeplorasian hanya terjadi pada sekitaran yang sempit disekeliling *centroid* (Cui, Potok, dan Palathingal, 2005; Alfina, Santosa, dan Barakbah, 2012). Solusi yang bersifat *local optimum* berdampak pada kualitas hasil pengklasteran menjadi kurang baik (Bhusare dan Bansode, 2014). Telah banyak penelitian yang membahas tentang solusi inisialisasi *centroid* algoritma *k-means*, antara lain Bhusare dan Bansode (2014), Ahmed dan Rahim (2017), dan semua penelitian ini bertujuan untuk mendapatkan inisialisasi *centroid* yang tepat.

Kondisi *local optimum* pada algoritma *k-means* terjadi karena pengeplorasian hanya pada sekitaran sempit disekeliling *centroid*. Maka dari itu perlu dicari kondisi yang dapat memenuhi kondisi *global optimum* dimana pengeplorasian terjadi untuk seluruh ruang pencarian (*search space*). Pencarian kondisi *global optimum* dapat dibantu dengan metode-metode yang mempunyai sifat pengoptimilisasian. Ada dua pendekatan dalam metode-metode pengoptimalisasian yaitu deterministik dan probabilistik (Suyanto, 2010).

Algoritma *Ant Colony Optimization* (ACO) merupakan sebuah turunan dari algoritma *Ant colony*. Dari semua varian *ant colony* yang ada, Dorigo menyimpulkan bahwa algoritma *Ant Colony Optimization* (ACO) merupakan yang terbaik dari varian yang ada (Dorigo, 2004). Pendapat tersebut dikuatkan oleh penelitian yang dilakukan oleh Leksono (2009) bahwa ACO memiliki hasil perhitungan yang relative mendekati optimum.

ACO merupakan suatu algoritma semut didalam membentuk suatu koloni. Algoritma ACO dapat menghindari dari permasalahan lokal optimal dan terbukti memiliki solusi global (Wicaksana dan Widiartha, 2012). Berdasarkan kehandalan metode ACO dalam menemukan solusi *global optimum* tersebut maka tulisan ini mencoba mengangkat Algoritma ACO sebagai acuan untuk dilakukan tindakan lanjutan yang bertujuan untuk menguji pengaruh metode ACO dalam penentuan *centroid* awal terhadap kualitas hasil pengklasteran algoritma *k-means*.

Selain itu, algoritma *k-means* juga mudah terjebak pada permasalahan *curse of dimentionality* pada data berdimensi tinggi. Pada data berdimensi tinggi kompleksitas perhitungan jarak akan meningkat secara eksponensial sebanyak

dimensi data (Saranya dan Krishnakumari, 2011) dan juga pada saat dimensi meningkat letak data akan menjadi semakin menjauh karena titik data berada di berbagai dimensi yang berakibat data dianggap memiliki jarak yang sama pada proses perhitungan jarak. Selain itu, data berdimensi tinggi biasanya hanya sebagian kecil dimensi yang relevan dengan kelompok tertentu, dimensi yang tidak relevan dapat menyebabkan banyak kebisingan (*noise*) dan menutupi kelompok nyata yang akan ditemukan. Kebanyakan algoritma pengklasteran bagus bekerja pada dimensi yang rendah (Hidayati, Ihsan, dan Danny, 2017). Untuk mengatasi permasalahan tersebut peran reduksi dimensi dibutuhkan. Menurut penelitian Hidayati, Ihsan, dan Danny (2017) dengan dilakukannya reduksi dimensi pada pengklasteran menggunakan metode *Singular Value Decomposition* (SVD) dapat meningkatkan akurasi hasil pengklasteran.

1.3 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah bagaimana pengaruh algoritma *Ant Colony Optimization* dalam penentuan *centroid* awal terhadap kualitas hasil pengklasteran *k-Means*

Untuk menjawab rumusan masalah tersebut, diuraikan beberapa *research question* sebagai berikut :

1. Bagaimana mekanisme kerja metode *Ant Colony Optimization* dan algoritma *k-means*?

2. Bagaimana metode *Ant Colony Optimization* dapat menentukan *centroid* awal algoritma *k-means*?
3. Bagaimana perbandingan hasil kualitas pengklasteran algoritma *k-means* dengan penentuan *centroid* awal menggunakan metode *Ant Colony Optimization* terhadap algoritma *k-means* dengan penentuan *centroid* awal secara acak?

1.4 Tujuan Penelitian

Tujuan dilakukannya penelitian adalah sebagai berikut :

1. Mengetahui mekanisme baru pengoptimalisasian algoritma *k-means* menggunakan metode ACO.
2. Mengetahui cara kerja metode ACO dalam menentukan *centroid* awal algoritma *k-means*.
3. Mengetahui perbandingan hasil kualitas pengklasteran algoritma *k-means* dengan penentuan *centroid* awal menggunakan metode ACO dan algoritma *k-means* dengan penentuan *centroid* awal secara acak.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut :

1. Memperoleh mekanisme baru pengoptimalisasian algoritma *k-means* menggunakan metode ACO.
2. Memahami cara kerja metode ACO dalam menentukan *centroid* awal algoritma *k-means*.

3. Memahami perbandingan kualitas hasil pengklasteran algoritma *k-means* dengan penentuan *centroid* awal menggunakan metode ACO dan algoritma *k-means* dengan penentuan *centroid* awal secara acak.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Pada penelitian ini metode pengklasteran yang diterapkan hanya metode pengklasteran *k-Means*.
2. Algoritma yang digunakan untuk menentukan titik pusat pada *k-means* adalah algoritma *Ant Colony Optimization*.
3. Data yang digunakan berupa data teks dalam bahasa Indonesia, yang diperoleh dari situs *garuda.ristekdikti.go.id*.
4. Metode yang digunakan untuk mereduksi dimensi adalah metode *Singular Value Decompositon (SVD)*.
5. Analisa perbandingan hasil pengklasteran menggunakan uji *Anova*.
6. Evaluasi hasil pengklasteran akan diukur dengan menghitung jarak antar *cluster* menggunakan metode *Davies-Bouldin Index*.

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 menjelaskandasar-dasar teori yang digunakan dalam penelitian. Pembahasan dalam bab ini meliputi hasil penelitian terdahulu yang relevan 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

Pemilihan dan penentuan *initial centroid* sangat berpengaruh pada pengklasteran *k-Means*. Dalam penentuan *initial centroid k-Means* dapat dilakukan dengan beberapa cara seperti dibangkitkan secara acak maupun dibangkitkan menggunakan beberapa metode lain, salah satunya yaitu algoritma *Ant Colony Optimization*. *Centroid* yang dibangkitkan secara acak sering kali menyebabkan *k-Means* terjebak pada *local optimum* dan menyebabkan kinerja hasil kluster masih kurang optimal. Untuk mengatasi masalah tersebut dapat dilakukan penentuan *centroid* menggunakan metode algoritma *Ant Colony Optimization* sebagai algoritma dengan teknik pencarian optimasi baru untuk menghindari *local minimal* dan menemukan *global optimum*.

DAFTAR PUSTAKA

- A. Leksono, “Algoritma *Ant Colony Optimization* (ACO) Untuk Menyelesaikan *Traveling Salesman Problem* (TSP),” Fakultas FMIPA Universitas Diponegoro, Semarang, 2009.
- Ahmed, T. and Rahim, M. S. (2017) ‘*An Initial Centroid Selection Method based on Radial and Angular Coordinates for K-means Algorithm*’, pp. 22–24.
- Alfina, T., Santosa, B. and Barakbah, R. (2012) ‘Analisa Perbandingan Metode *Hierarchical Clustering* , *K-means* dan Gabungan Keduanya dalam *Cluster Data* (Studi kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS)’, 1, p. A-521.
- Bhusare, B. B. and Bansode, S. M. (2014) ‘*Centroids Initialization for K-Means Clustering using Improved Pillar Algorithm*’, 3(4), pp. 1317–1322.
- Bide, P. and Shedge, R. (2015) ‘*Improved Document Clustering using k-means algorithm*’, Proceedings of 2015 IEEE International Conference on Electrical, Computer and Communication Technologies, ICECCT 2015, (978), pp. 0–4. doi: 10.1109/ICECCT.2015.7226065.
- Cui, X., Potok, T. E. and Palathingal, P. (2005) ‘*Document Clustering using Particle Swarm Optimization*’, pp. 1–7.
- Dorigo, M., dan Gambardela, L., (1996). *Ant Colony System: A Cooperative learning Approach to the Traveling Salesman Problem*. Tech.Rep/IRIDIA/1965-005, Universite Libre de Bruxelles, Belgium.
- Eka Merliana, N. P., Ernawati and Santoso, A. J. (2015) ‘Analisa Penentuan Jumlah *Cluster* Terbaik Pada Metode *K-Means*’, *Prosiding Seminar Nasional Multi Disiplin Ilmu&Call for Papers Unisbank (Sendi_U)*, pp. 978–979.
- Han, J., Kamber, M. and Pei, J. (2012) *Datamining Concepts and Techniques*.

third. Morgan Kaufmann.

- Herwijayanti, B., Ratnawati, D. E. and Muflikhah, L. (2018) 'Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan *Cosine Similarity*', 2(1), pp. 306–312.
- Hidayati, N., Ihsan, M. and Danny, M. (2017) 'PENGARUH *SINGULAR VALUE DECOMPOSITION* TERHADAP METODE – METODE CLUSTERING', pp. 95–104.
- Kadhim, A. I., Cheah, Y. and Ahamed, N. H. (2014) '*Text Document Preprocessing and Dimension Reduction Techniques for Text Document Clustering*'. doi: 10.1109/ICAIET.2014.21.
- Mishra, R. K., Saini, K. and Bagri, S. (2015) '*Text document clustering on the basis of inter passage approach by using K-means*', *International Conference on Computing, Communication and Automation, ICCCA 2015*, pp. 110–113. doi: 10.1109/CCAA.2015.7148354.
- Murti, D. H., Suciati, N. and Nanjaya, D. J. (2014) '*Clustering Data Non-Numerik Dengan Pendekatan Algoritma K-Means dan Hamming Distance Studi Kasus Biro Jodoh*', *JUTI: Jurnal Ilmiah Teknologi Informasi*, 4(1), p. 46. doi: 10.12962/j24068535.v4i1.a245.
- P, B., Sunil, K. P. V and Balachandran, K. P. (2013) '*K-Means Clustering - Review of Various Methods for Initial Selection of Centroids*', 4(8), p. 4.
- Prasetyo, E. (2012) *Data Mining - Konsep dan Aplikasi Menggunakan MATLAB*. Edited by N. Wk. C.V ANDI OFFSET.
- Rahmawati, L., Cahyani, A. D. and Putro, S. S. (2015) 'PEMANFAATAN METODE *CLUSTER SOM – IDB* SEBAGAI ANALISA PENGELOMPOKAN PENERIMA BEASISWA', (11–18).
- S. T. Marco Dorigo, *Ant Colony Optimization*, London England and Cambridge, Massachussets: A Bradford Book and The MIT Press, 2004.

- Santosa, B. and Umam, A. (2018) *Data Mining dan Big Data Analytics*. pertama. Edited by Isa. Penebar Media Pustaka.
- Santoso, B. (2007) *data mining: teknik pemanfaatan data untuk keperluan bisnis*. first. indonesia: graha ilmu.
- Saranya, M. and Krishnakumari, P. (2011) ‘*An Efficient Algorithm to fix Initial Centroid for Clustering High Dimensional data*’, 2(4), pp. 595–598.
- Suyanto (2010) *ALGORITMA OPTIMASI (Deterministik atau Probabilistik)*. pertama. graha ilmu.
- Wibawa, S. M. and Novianti, K. D. P. (2017) ‘*Reduksi Fitur Untuk Optimalisasi Klasifikasi Tumor Payudara Berdasarkan Data Citra FNA*’, pp. 73–78.
- Wicaksana, M. K., Widiartha, M. K. (2012) ‘*PENERAPAN METODE ANT COLONY OPTIMIZATION PADA METODE K-HARMONIC MEANS UNTUK KLASTERISASI DATA*’, pp. hh.
- Widiarina and Wahono, R. S. (2015) ‘*Algoritma Cluster Dinamik untuk Optimasi Cluster pada Algoritma K-Means dalam Pemetaan Nasabah Potensial*’, 1(1), pp. 33–36.