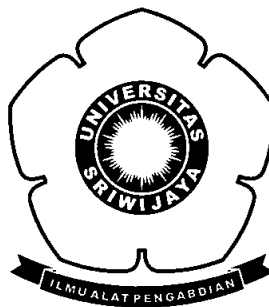


**PENGARUH METODE *PARTICLE SWARM OPTIMIZATION*
DALAM PENENTUAN *CENTROID* AWAL TERHADAP
KUALITAS HASIL *CLUSTERING* ALGORITMA *K-MEANS***

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Studi Strata-1 Pada
Jurusan Teknik Informatika Reguler Fakultas Ilmu Komputer



Oleh :

Muhammad Syazili

09021181520135

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

LEMBAR PENGESAHAN TUGAS AKHIR

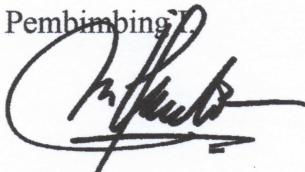
**PENGARUH METODE *PARTICLE SWARM OPTIMIZATION*
DALAM PENENTUAN *CENTROID* AWAL TERHADAP
KUALITAS HASIL *CLUSTERING* ALGORITMA *K-MEANS***

Oleh :

MUHAMMAD SYAZILI
NIM : 09021181520135

Mengetahui,

Pembimbing I,

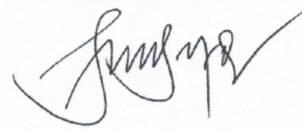


Ir. M. Ihsan Jambak, M.Sc.

NIP. 196804052013081201

Indralaya, 12 Agustus 2020

Pembimbing II,



Rusdi Effendi, M.Kom.

NIP. 198201022011021201

Mengetahui,

Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T.

NIP. 197706012009121004

CERTIFICATE OF PASSING

On Tuesday, July 28th, 2020, a final project examination session was held by Departement of Informatics Engineering, Faculty of Computer Sciene, Sriwijaya University.

Name : Muhammad Syazili
NIM : 09021181520135
Title : Effect Of Particle Swarm Optimization Method To Decide Early Centroid For Clustering Results Using K-Means

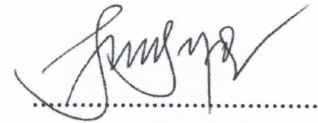
1. Supervisor I

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



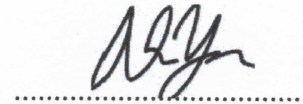
2. Supervisor II

Rusdi Effendi, M.Kom.
NIP. 198201022011021201



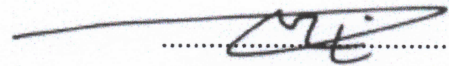
3. Examiner I

Novi Yusliani, M.T.
NIP.198211082012122001



4. Examiner II

Osvari Arsalan, S.Kom., M.T.
NIP. 198806282016011901



Approved,
Chariman of Informatics Engineering Departement,



Rifkie Primartha, M.T.
NIP. 197706012009121004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Muhammad Syazili

NIM : 09021181520135

Program Studi : Teknik Informatika

Judul Skripsi : Pengaruh Metode *Particle Swarm Optimization* Dalam Penentuan *Centroid* Awal Terhadap Kualitas Hasil *Clustering* Algoritma *K-Means*

Hasil Pemeriksaan *Software iThenticate / Turnitin* : 10%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Jika ditemukan unsur penjiplakan/plagiat dalam laporan tugas akhir 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



Muhammad Syazili

NIM. 09021181520135

HALAMAN MOTTO DAN PERSEMBAHAN

Motto :

“khairunnas anfa’uhum linnas”

sebaik-baik manusia diantaramu adalah yang paling banyak memberi manfaat bagi orang lain

Kupersembahkan karya tulis ini kepada :

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

EFFECT OF PARTICLE SWARM OPTIMIZATION METHOD TO DECIDE
EARLY CENTROID FOR CLUSTERING RESULTS USING K-MEANS

by:

Muhammad Syazili
09021181520135

ABSTRACT

The quality of clustering results using k-means depend on initialization of early centroid. Initialization of early centroid generated randomly, produce in a convergen condition such a local optimum, therefore it needs to be found that k-means algorithm be able to produce in a convergen condition such a global optimum using Particle Swarm Optimization. In addition, most of clustering algorithm works well in handling low dimentional data and low dimentions can be achieved by doing dimentional reduction. From this research, the determination of early centroid using Particle Swarm Optimization can improve the quality of clustering results using k-means compared to the determination of early centroid randomly without dimentional reduction by showing a significant decrease in the DBI value of 31,61818134108936%. Meanwhile, the determination of early centroid using Particle Swarm Optimization can improve the quality of clustering results using k-means compared to determination of early centroid radomly with dimentional reduction by showing significant decrease in the DBI value of 38,5403727707036%.

Keywords: K-Means Algorithm, Determination of Early Centroid Using K-Means, Particle Swarm Optimization, Dimentional Reduction.

PENGARUH METODE *PARTICLE SWARM OPTIMIZATION* DALAM
PENENTUAN *CENTROID* AWAL TERHADAP KUALITAS HASIL
CLUSTERING ALGORITMA *K-MEANS*

Oleh:

Muhammad Syazili
09021181520135

ABSTRAK

Kualitas hasil *clustering* algoritma *k-means* sangat tergantung pada inisialisasi *centroid* awal. Inisialisasi *centroid* awal yang dibangkitkan secara acak sering kali menghasilkan kondisi konvergensi yang bersifat *local optimum*, maka dari itu perlu dicarikan cara agar algoritma *k-means* memenuhi kondisi konvergensi yang bersifat *global optimum* yaitu salah satunya menggunakan metode *Particle Swarm Optimization* (PSO). Selain itu, kebanyakan algoritma *clustering* baik dalam menangani data dimensi rendah, dimensi rendah dapat dicapai dengan cara melakukan reduksi dimensi. Dari hasil penelitian didapat bahwa dengan penentuan *centroid* awal menggunakan metode *Particle Swarm Optimization* dapat meningkatkan kualitas hasil *clustering* algoritma *k-means* dibandingkan penentuan *centroid* awal ditetapkan secara acak pada data tanpa dilakukan reduksi dimensi yaitu dengan ditunjukkan penurunan nilai DBI yang signifikan sebesar 31,61818134108936%. Sedangkan penentuan *centroid* awal menggunakan metode *Particle Swarm Optimization* dapat meningkatkan kualitas hasil *clustering* algoritma *k-means* dibandingkan penentuan *centroid* awal ditetapkan secara acak pada data dengan dilakukan reduksi dimensi yaitu dengan ditunjukkan penurunan nilai DBI yang signifikan sebesar 38,5403727707036%.

Kata Kunci: Algoritma *K-Means*, Penentuan *Centroid* Awal Algoritma *K-Means*, *Particle Swarm Optimization*, Reduksi Dimensi.

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. Ibuku, Ayahku, dan seluruh keluarga besarku yang selalu mendokan serta memberikan dukungan kepadaku.
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. Ibu Alvi Syahrini Utami, M.Kom. 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 dan Bapak Rusdi Effendi, M.Kom. selaku dosen pembimbing II yang telah membimbing, mengarahkan dan memberikan motivasi Penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.

6. Ibu Novi Yusliani, M.T. selaku penguji I dan Bapak Osvari Arsalan, 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. Mbak Winda, Mbak Wiwin, Kak Hafez, 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

Muhammad Syazili

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN TUGAS AKHIR	ii
TANDA LULUS SIDANG TUGAS AKHIR.....	iii
HALAMAN PERNYATAAN	iv
HALAMAN MOTTO DAN PERSEMBAHAN.....	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xv
DAFTAR GAMBAR	xxiv
BAB I PENDAHULUAN.....	I-1.
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-6.
1.6 Batasan Masalah	I-7.
1.7 Sistematika Penulisan	I-7.
1.8 Kesimpulan.....	I-9.
BAB II KAJIAN LITERATUR	II-1.
2.1 Pendahuluan.....	II-1.
2.2 Landasan Teori	II-1.

2.2.1 Pengklasteran.....	II-1.
2.2.2 Algoritma <i>K-means</i>	II-2.
2.2.3 Konversi Data	II-5.
2.2.4 Reduksi Dimensi.....	II-7.
2.2.4.1 Metode <i>Singular Value Decomposition</i>	II-8.
2.2.5 Metode <i>Particle Swarm Optimization</i>	II-10.
2.2.6 Metode <i>Davies Bouldin Index (DBI)</i>	II-16.
2.2.7 <i>Rational Unified Process</i>	II-17.
2.3 Penelitian Lain yang Relevan.....	II-19.
2.4 Kesimpulan.....	II-22.
BAB III METODOLOGI PENELITIAN.....	III-1.
3.1 Pendahuluan.....	III-1.
3.2 Unit Penelitian dan Data.....	III-1.
3.3 Tahapan Penelitian	III-1.
3.3.1 Menetapkan Kerangka Kerja	III-2.
3.3.2 Menetapkan Kriteria Pengujian	III-10.
3.3.3 Menetapkan Format Data Pengujian.....	III-11.
3.3.3.1 Penetapan Jumlah Kluster Optimum.....	III-11.
3.3.3.2 Pengujian Parameter Metode PSO dalam Penentuan Centroid Awal Algoritma <i>K-Means</i>	III-12.
3.3.3.3 Pengujian Hasil Pengklasteran Algoritma <i>K-Means</i>	III-15.
3.3.4 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian.	III-17.
3.3.5 Melakukan Pengujian Penelitian	III-17.
3.3.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan Penelitian	III-18.
3.4 Metode Pengembangan Perangkat Lunak	III-18.
3.4.1 <i>Rational Unified Process (RUP)</i>	III-18.
3.4.1.1 Fase Insepsi	III-19.
3.4.1.2 Fase Elaborasi	III-19.
3.4.1.3 Fase Konstruksi.....	III-20.
3.4.1.4 Fase Transisi.....	III-20.
3.5 Manajemen Proyek Penelitian	III-21.

3.6 Kesimpulan.....	III-31.
BAB IV PENGEMBANGAN PERANGKAT LUNAK	IV-1.
4.1 Pendahuluan	IV-1.
4.2 Fase Insepsi	IV-1.
4.2.1 <i>Business Modeling</i>	IV-2.
4.2.2 <i>Requirements</i>	IV-6.
4.2.3 <i>Analysis and Design</i>	IV-11.
4.3 Fase Elaborasi	IV-13.
4.3.1 <i>Business Modeling</i>	IV-13.
4.3.2 <i>Requirements</i>	IV-19.
4.3.3 <i>Analysis and Design</i>	IV-25.
4.4 Fase Konstruksi	IV-28.
4.4.1 <i>Business Modeling</i>	IV-29.
4.4.2 <i>Requirements</i>	IV-31.
4.4.3 <i>Implementation</i>	IV-31.
4.4.3.1 Implementasi Kelas.....	IV-31.
4.4.4 <i>Testing</i>	IV-33.
4.5 Fase Transisi	IV-39.
4.5.1 <i>Business Modeling</i>	IV-39.
4.5.2 <i>Requirements</i>	IV-40.
4.5.3 <i>Testing</i>	IV-40.
4.6 Kesimpulan	IV-55.
BAB V HASIL DAN ANALISIS PENELITIAN	V-1.
5.1 Pendahuluan	V-1.
5.2 Konfigurasi Percobaan	V-2.
5.3 Hasil Pengujian	V-3.
5.3.1 Hasil Pengujian Jumlah Kluster Optimum.....	V-3.
5.3.2 Hasil Pengujian Pengklasteran Algoritma <i>K-Means</i> dengan Penentuan Pusat Kluster Awal Ditetapkan Secara Acak	V-4.

5.3.3 Hasil Pengujian Parameter Metode <i>Particle Swarm Optimization</i>	V-6.
5.3.3.1 Hasil Pengujian Jumlah Iterasi Metode <i>Particle Swarm Optimization</i>	V-6.
5.3.3.2 Hasil Pengujian Jumlah Partikel Metode <i>Particle Swarm Optimization</i>	V-9.
5.3.3.3 Hasil Pengujian <i>Learning Rate</i> 1 Metode <i>Particle Swarm Optimization</i>	V-12.
5.3.3.4 Hasil Pengujian <i>Learning Rate</i> 2 Metode <i>Particle Swarm Optimization</i>	V-15.
5.3.4 Hasil Pengujian Pengklasteran Algoritma <i>K-Means</i> dengan Penentuan Pusat Kluster Awal Menggunakan Metode <i>Particle Swarm Optimization</i>	V-18.
5.4 Olah Data.....	V-20.
5.4.1 Olah Data Hasil Pengujian Jumlah Kluster Optimum.....	V-20.
5.4.2 Olah Data Hasil Perbandingan Nilai DBI	V-21.
5.4.3 Olah Data Hasil Perbandingan Waktu Komputasi	V-30.
5.4.4 Olah Data Hasil Perbandingan Jumlah Iterasi.....	V-40.
5.5 Analisis Hasil Pengujian	V-50.
5.5.1 Analisis Hasil Pengujian Nilai Jumlah Kluster Optimum..	V-50.
5.5.2 Analisis Hasil Pengujian Parameter Metode <i>Particle Swarm Optimization</i>	V-51.
5.5.3 Analisis Hasil Pengujian Pengklasteran Algoritma <i>K-Means</i> dengan Penentuan Pusat Kluster Awal Ditetapkan Secara Acak pada Data dengan Dilakukan Reduksi Dimensi dan tanpa Dilakukan Reduksi Dimensi.....	V-52.
5.5.4 Analisis Hasil Pengujian Pengklasteran Algoritma <i>K-Means</i> dengan Penentuan Pusat Kluster Awal Menggunakan Metode <i>Particle Swarm Optimization</i> pada Data dengan Dilakukan Reduksi Dimensi dan tanpa Dilakukan Reduksi Dimensi	V-53.

5.5.5 Analisis Hasil Pengujian Pengklasteran Algoritma <i>K-Means</i> dengan Penentuan Pusat Klaster Awal Ditetapkan Secara Acak dengan Pengklasteran Algoritma <i>K-Means</i> dengan Penentuan Pusat Klaster Awal Menggunakan Metode <i>Particle Swarm Optimization</i> pada Data tanpa Dilakukan Reduksi Dimensi	V-54.
5.5.6 Analisis Hasil Pengujian Pengklasteran Algoritma <i>K-Means</i> dengan Penentuan Pusat Klaster Awal Ditetapkan Secara Acak dengan Pengklasteran Algoritma <i>K-Means</i> dengan Penentuan Pusat Klaster Awal Menggunakan Metode <i>Particle Swarm Optimization</i> pada Data dengan Dilakukan Reduksi Dimensi	V-56.
5.6 Kesimpulan.....	V-57.
BAB VI KESIMPULAN DAN SARAN	VI-1.
6.1 Pendahuluan	VI-1.
6.2 Kesimpulan.....	VI-1.
6.3 Saran	VI-2.
DAFTAR PUSTAKA	xxix
LAMPIRAN I KODE PROGRAM	xxxiii

DAFTAR TABEL

	Halaman
Tabel III – 1 Format Data Pengujian Penetapan Jumlah Kluster Optimum....	III-12
Tabel III – 2 Format Data Pengujian Parameter Jumlah Iterasi PSO	III-12
Tabel III – 3 Format Data Pengujian Parameter Jumlah Partikel PSO	III-13
Tabel III – 4 Format Data Pengujian Parameter <i>Learning Rate</i> 1 Metode PSO	III-14
Tabel III – 5 Format Data Pengujian Parameter <i>Learning Rate</i> 2 Metode PSO	III-15
Tabel III – 6 Format Data Pengujian Hasil Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat kluster Awal Ditetapkan Secara Acak.....	III-16
Tabel III – 7 Format Data Pengujian Hasil Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat kluster Awal Ditetapkan Metode PSO	III-16
Tabel III – 8 Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure</i>	III-21
Tabel IV – 1 Kebutuhan Fungsional	IV-3
Tabel IV – 2 Tabel Kebutuhan Non-Fungsional.....	IV-3
Tabel IV – 3 Definisi Aktor	IV-5
Tabel IV – 4 Definisi <i>Use Case</i>	IV-5
Tabel IV – 5 Tabel Implementasi Kelas	IV-31
Tabel IV – 6 Rencana Pengujian Konversi Data	IV-33
Tabel IV – 7 Rencana Pengujian Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat Kluster Awal Ditetapkan Secara Acak tanpa	

Dilakukan Reduksi Dimensi	IV-34
Tabel IV – 8 Rencana Pengujian Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat Klaster Awal Ditetapkan Menggunakan Metode PSO tanpa Dilakukan Reduksi Dimensi	IV-35
Tabel IV – 9 Rencana Pengujian Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat Klaster Awal Ditetapkan Secara Acak dengan Dilakukan Reduksi Dimensi.....	IV-37
Tabel IV – 10 Rencana Pengujian Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat Klaster Awal Ditetapkan Menggunakan Metode PSO dengan Dilakukan Reduksi Dimensi	IV-38
Tabel IV – 11 Implementasi Pengujian Konversi Data	IV-41
Tabel IV – 12 Implementasi Pengujian Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat Klaster Awal Ditetapkan Secara Acak tanpa Dilakukan Reduksi Dimensi	IV-43
Tabel IV – 13 Implementasi Pengujian Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat Klaster Awal Ditetapkan Menggunakan Metode PSO tanpa Dilakukan Reduksi Dimensi	IV-46
Tabel IV – 14 Implementasi Pengujian Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat Klaster Awal Ditetapkan Secara Acak dengan Dilakukan Reduksi Dimensi.....	IV-49
Tabel IV – 15 Implementasi Pengujian Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat Klaster Awal Ditetapkan Menggunakan Metode PSO dengan Dilakukan Reduksi Dimensi	IV-52

Tabel V – 1 Hasil Pengujian Jumlah Kluster Optimum.....	V-4
Tabel V – 2 Hasil Pengujian Pengklasteran Algoritma <i>K-Means</i> dengan Penentuan Pusat Kluster Awal Ditetapkan Secara Acak	V-5
Tabel V – 3 Hasil Pengujian Jumlah Iterasi Metode <i>Particle Swarm Optimization</i>	V-7
Tabel V – 4 Hasil Pengujian Jumlah Partikel Metode <i>Particle Swarm Optimization</i>	V-10
Tabel V – 5 Hasil Pengujian Nilai <i>Learning Rate 1</i> Metode <i>Particle Swarm Optimization</i>	V-13
Tabel V – 6 Hasil Pengujian <i>Learning Rate 2</i> Metode <i>Particle Swarm Optimization</i>	V-16
Tabel V – 7 Hasil Pengujian Pengklasteran Algoritma <i>K-Means</i> dengan Penentuan Pusat Kluster Awal Menggunakan Metode <i>Particle Swarm Optimization</i>	V-19
Tabel V – 8 Nilai Rata-Rata DBI Setiap Metode Pengklasteran	V-21
Tabel V – 9 Hasil Uji Normalitas Perbandingan Nilai DBI.....	V-22
Tabel V – 10 Hasil Peringkat Rata-Rata Perbandingan Nilai DBI Antara Metode Pengklastran Menggunakan Uji <i>Kruskal-Wallis</i>	V-23
Tabel V – 11 Hasil Tes Statistik Perbandingan Nilai DBI Antara Metode Pengklastran Menggunakan Uji <i>Kruskal-Wallis</i>	V-23
Tabel V – 12 Hasil Peringkat Rata-Rata untuk Perbandingan Nilai DBI pada Metode Pengklasteran <i>K-Means</i> dan <i>K-Means+PSO</i> Menggunakan	

Uji <i>Mann-Whitney</i>	V-24
Tabel V – 13 Hasil Tes Statistik untuk Perbandingan Nilai DBI pada Metode Pengklasteran K-Means dan K-Means+PSO Menggunakan Uji <i>Mann-Whitney</i>	V-24
Tabel V – 14 Hasil Peringkat Rata-Rata untuk Perbandingan Nilai DBI pada Metode Pengklasteran K-Means dan K-Means+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-25
Tabel V – 15 Hasil Tes Statistik untuk Perbandingan Nilai DBI pada Metode Pengklasteran K-Means dan K-Means+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-25
Tabel V – 16 Hasil Peringkat Rata-Rata untuk Perbandingan Nilai DBI pada Metode Pengklasteran K-Means dan K-Means+PSO+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-26
Tabel V – 17 Hasil Tes Statistik untuk Perbandingan Nilai DBI pada Metode Pengklasteran K-Means dan K-Means+PSO+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-26
Tabel V – 18 Hasil Peringkat Rata-Rata untuk Perbandingan Nilai DBI pada Metode Pengklasteran K-Means+SVD dan K-Means+PSO Menggunakan Uji <i>Mann-Whitney</i>	V-27
Tabel V – 19 Hasil Tes Statistik untuk Perbandingan Nilai DBI pada Metode Pengklasteran K-Means+SVD dan K-Means+PSO Menggunakan Uji <i>Mann-Whitney</i>	V-27
Tabel V – 20 Hasil Peringkat Rata-Rata untuk Perbandingan Nilai DBI	

pada Metode Pengklasteran K-Means+SVD dan K-Means+PSO+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-28
Tabel V – 21 Hasil Tes Statistik untuk Perbandingan Nilai DBI pada Metode Pengklasteran K-Means+SVD dan K-Means+PSO+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-29
Tabel V – 22 Hasil Peringkat Rata-Rata untuk Perbandingan Nilai DBI pada Metode Pengklasteran K-Means+PSO dan K-Means+PSO+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-29
Tabel V – 23 Hasil Tes Statistik untuk Perbandingan Nilai DBI pada Metode Pengklasteran K-Means+PSO dan K-Means+PSO+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-30
Tabel V – 24 Nilai Rata-Rata Waktu Komputasi Setiap Metode Pengklasteran	V-31
Tabel V – 25 Hasil Uji Normalitas Perbandingan Waktu Komputasi	V-32
Tabel V – 26 Hasil Peringkat Rata-Rata Perbandingan Waktu Komputasi Antara Metode Pengklastran Menggunakan Uji <i>Kruskal-Wallis</i>	V-32
Tabel V – 27 Hasil Tes Statistik Perbandingan Waktu Komputasi Antara Metode Pengklastran Menggunakan Uji <i>Kruskal-Wallis</i>	V-33
Tabel V – 28 Hasil Peringkat Rata-Rata untuk Perbandingan Waktu Komputasi pada Metode Pengklasteran K-Means dan K-Means+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-34
Tabel V – 29 Hasil Tes Statistik untuk Perbandingan Waktu Komputasi pada Metode Pengklasteran K-Means dan K-Means+SVD Menggunakan	

Uji <i>Mann-Whitney</i>	V-34
Tabel V – 30 Hasil Peringkat Rata-Rata untuk Perbandingan Waktu Komputasi pada Metode Pengklasteran K-Means dan K-Means+PSO Menggunakan Uji <i>Mann-Whitney</i>	
	V-35
Tabel V – 31 Hasil Tes Statistik untuk Perbandingan Waktu Komputasi pada Metode Pengklasteran K-Means dan K-Means+PSO Menggunakan Uji <i>Mann-Whitney</i>	
	V-35
Tabel V – 32 Hasil Peringkat Rata-Rata untuk Perbandingan Waktu Komputasi pada Metode Pengklasteran K-Means dan K-Means+PSO+SVD Menggunakan Uji <i>Mann-Whitney</i>	
	V-36
Tabel V – 33 Hasil Tes Statistik untuk Perbandingan Waktu Komputasi pada Metode Pengklasteran K-Means dan K-Means+PSO+SVD Menggunakan Uji <i>Mann-Whitney</i>	
	V-36
Tabel V – 34 Hasil Peringkat Rata-Rata untuk Perbandingan Waktu Komputasi pada Metode Pengklasteran K-Means+SVD dan K-Means+PSO Menggunakan Uji <i>Mann-Whitney</i>	
	V-37
Tabel V – 35 Hasil Tes Statistik untuk Perbandingan Waktu Komputasi pada Metode Pengklasteran K-Means+SVD dan K-Means+PSO Menggunakan Uji <i>Mann-Whitney</i>	
	V-37
Tabel V – 36 Hasil Peringkat Rata-Rata untuk Perbandingan Waktu Komputasi pada Metode Pengklasteran K-Means+SVD dan K-Means+PSO+SVD Menggunakan Uji <i>Mann-Whitney</i>	
	V-38
Tabel V – 37 Hasil Tes Statistik untuk Perbandingan Waktu Komputasi	

pada Metode Pengklasteran K-Means+SVD dan K-Means+PSO+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-38
Tabel V – 38 Hasil Peringkat Rata-Rata untuk Perbandingan Waktu Komputasi pada Metode Pengklasteran K-Means+PSO dan K-Means+PSO+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-39
Tabel V – 39 Hasil Tes Statistik untuk Perbandingan Waktu Komputasi pada Metode Pengklasteran K-Means+PSO dan K-Means+PSO+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-40
Tabel V – 40 Nilai Rata-Rata Jumlah Iterasi Setiap Metode Pengklasteran.....	V-41
Tabel V – 41 Hasil Uji Normalitas Perbandingan Jumlah Iterasi.....	V-42
Tabel V – 42 Hasil Peringkat Rata-Rata Perbandingan Jumlah Iterasi Antara Metode Pengklastran Menggunakan Uji <i>Kruskal-Wallis</i>	V-42
Tabel V – 43 Hasil Tes Statistik Perbandingan Jumlah Iterasi Antara Metode Pengklastran Menggunakan Uji <i>Kruskal-Wallis</i>	V-43
Tabel V – 44 Hasil Peringkat Rata-Rata untuk Perbandingan Jumlah Iterasi pada Metode Pengklasteran K-Means dan K-Means+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-44
Tabel V – 45 Hasil Tes Statistik untuk Perbandingan Jumlah Iterasi pada Metode Pengklasteran K-Means dan K-Means+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-44
Tabel V – 46 Hasil Peringkat Rata-Rata untuk Perbandingan Jumlah Iterasi pada Metode Pengklasteran K-Means dan K-Means+PSO Menggunakan Uji <i>Mann-Whitney</i>	V-45

Tabel V – 47 Hasil Tes Statistik untuk Perbandingan Jumlah Iterasi pada Metode Pengklasteran K-Means dan K-Means+PSO Menggunakan Uji <i>Mann-Whitney</i>	V-45
Tabel V – 48 Hasil Peringkat Rata-Rata untuk Perbandingan Jumlah Iterasi pada Metode Pengklasteran K-Means dan K-Means+PSO+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-46
Tabel V – 49 Hasil Tes Statistik untuk Perbandingan Jumlah Iterasi pada Metode Pengklasteran K-Means dan K-Means+PSO+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-46
Tabel V – 50 Hasil Peringkat Rata-Rata untuk Perbandingan Jumlah Iterasi pada Metode Pengklasteran K-Means+SVD dan K-Means+PSO Menggunakan Uji <i>Mann-Whitney</i>	V-47
Tabel V – 51 Hasil Tes Statistik untuk Perbandingan Jumlah Iterasi pada Metode Pengklasteran K-Means+SVD dan K-Means+PSO Menggunakan Uji <i>Mann-Whitney</i>	V-47
Tabel V – 52 Hasil Peringkat Rata-Rata untuk Perbandingan Jumlah Iterasi pada Metode Pengklasteran K-Means+SVD dan K-Means+PSO+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-48
Tabel V – 53 Hasil Tes Statistik untuk Perbandingan Jumlah Iterasi pada Metode Pengklasteran K-Means+SVD dan K-Means+PSO+SVD Menggunakan Uji <i>Mann-Whitney</i>	V-48
Tabel V – 54 Hasil Peringkat Rata-Rata untuk Perbandingan Jumlah Iterasi pada Metode Pengklasteran K-Means+PSO dan	

K-Means+PSO+SVD Menggunakan Uji *Mann-Whitney* V-49

Tabel V – 55 Hasil Tes Statistik untuk Perbandingan Jumlah Iterasi pada

Metode Pengklasteran K-Means+PSO dan K-Means+PSO+SVD Menggunakan

Uji *Mann-Whitney*..... V-49

DAFTAR GAMBAR

	Halaman
Gambar II – 1 <i>Flowchart</i> Algoritma <i>K-means</i>	II – 4
Gambar II – 2 <i>Flowchart</i> Konversi Data	II – 6
Gambar II – 3 Matriks Dekomposisi <i>Singular Value Decomposition</i>	II – 8
Gambar II – 4 <i>Flowchart</i> Metode SVD dalam Melakukan Reduksi Dimensi	II – 9
Gambar II-5 <i>Flowchart</i> Metode PSO dalam Menentukan Pusat Klaster Awal Algoritma <i>K-Means</i>	II – 15
Gambar II – 6. Arsitektur <i>Rational Unified Process</i> (RUP)	II – 18
Gambar III-1. Tahapan Penelitian	III – 2
Gambar III-2. Kerangka Kerja Penelitian	III – 3
Gambar III-3. Penjadwalan Penelitian Tahap Menentukan Masalah dan Sistematika Penulisan Penelitian	III– 26
Gambar III-4. Panjadwalan Penelitian Tahap Menentukan Dasar Teori dan Penelitian Lain yang Relevan	III – 27
Gambar III-5. Penjadwalan Penelitian Tahap Menentukan Metodologi Penelitian	III – 27
Gambar III-6. Penjadwalan Penelitian Tahap Menentukan Metode Pengembangan Perangkat Lunak Fase <i>Inception</i>	III – 28
Gambar III-7. Penjadwalan Penelitian Tahap Menentukan Metode Pengembangan Perangkat Lunak Fase <i>Elaboration</i>	III – 29

Gambar III–8. Penjadwalan Penelitian Tahap Menentukan Metode Pengembangan Perangkat Lunak Fase <i>Construction</i>	III – 29
Gambar III–9. Penjadwalan Penelitian Tahap Menentukan Metode Pengembangan Perangkat Lunak Fase <i>Transition</i>	III – 30
Gambar III–10. Penjadwalan Penelitian Tahap Membuat Hasil, Analisa dan Kesimpulan Penelitian	III – 30
Gambar IV–1. Diagram <i>Use Case</i>	IV – 4
Gambar IV–2. Diagram Aktivitas Konversi Data	IV – 7
Gambar IV–3. Diagram Pengklasteran Menggunakan Algoritma K-Means dengan Pusat Klaster Awal Ditetapkan Secara Acak Tanpa Dilakukan Reduksi Dimensi	IV – 8
Gambar IV–4. Diagram Aktivitas Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat klaster Awal Ditetapkan Menggunakan Metode PSO tanpa Dilakukan Reduksi Dimensi	IV – 9
Gambar IV–5. Diagram Aktivitas Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat klaster Awal Ditetapkan Secara Acak dengan Dilakukan Reduksi Dimensi	IV – 10
Gambar IV–6. Diagram Aktivitas Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat klaster Awal Ditetapkan Menggunakan Metode PSO dengan Dilakukan Reduksi Dimensi.....	IV – 11
Gambar IV–7. Diagram <i>Sequence</i> Konversi Data	IV – 14
Gambar IV–8. Diagram <i>Sequence</i> Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat Klaster Awal Ditetapkan Secara Acak Tanpa	

Dilakukan Reduksi Dimensi	IV – 15
Gambar IV–9. Diagram <i>Sequence</i> Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat klaster Awal Ditetapkan Menggunakan Metode PSO tanpa Dilakukan Reduksi Dimensi	IV – 16
Gambar IV–10. Diagram <i>Sequence</i> Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat klaster Awal Ditetapkan Secara Acak dengan Dilakukan Reduksi Dimensi	IV – 17
Gambar IV–11. Diagram <i>Sequence</i> Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat klaster Awal Ditetapkan Menggunakan Metode PSO dengan Dilakukan Reduksi Dimensi	IV – 18
Gambar IV–12. <i>Mockup</i> Perangkat Lunak.....	IV – 19
Gambar IV–13. <i>Mockup</i> Bagian Memasukkan Data	IV – 19
Gambar IV–14. <i>Mockup</i> Bagian Menampilkan Jumlah Dokumen dan Jumlah <i>Term</i>	IV – 20
Gambar IV–15. <i>Mockup</i> Bagian Memilih Metode Pengklasteran.....	IV – 21
Gambar IV–16. <i>Mockup</i> Bagian Memasukkan Parameter Algoritma <i>K-means</i>	IV – 22
Gambar IV–17. <i>Mockup</i> Bagian Memasukkan Parameter Metode PSO	IV – 22
Gambar IV–18. <i>Mockup</i> Bagian Menjalankan Proses Pengklasteran.....	IV – 23
Gambar IV–19. <i>Mockup</i> Bagian Menampilkan Pusat klaster Akhir.....	IV – 23
Gambar IV–20. <i>Mockup</i> Bagian Menampilkan Hasil Pengklasteran	IV – 24
Gambar IV–21. <i>Mockup</i> Bagian Menampilkan Hasil Perhitungan Nilai DBI, Waktu Komputasi dan Jumlah Iterasi	IV – 25

Gambar IV–22. Diagram Aktivitas Konversi Data	IV – 26
Gambar IV–23. Diagram Aktivitas Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat klaster Awal Ditetapkan Secara Acak tanpa Dilakukan Reduksi Dimensi	IV – 26
Gambar IV–24. Diagram Aktivitas Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat klaster Awal Ditetapkan Menggunakan Metode PSO tanpa Dilakukan Reduksi Dimensi	IV – 27
Gambar IV–25. Diagram Aktivitas Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat klaster Awal Ditetapkan Secara Acak dengan Dilakukan Reduksi Dimensi	IV – 27
Gambar IV–26. Diagram Aktivitas Pengklasteran Menggunakan Algoritma <i>K-means</i> dengan Pusat klaster Awal Ditetapkan Menggunakan Metode PSO dengan Dilakukan Reduksi Dimensi	IV – 28
Gambar IV–27. Diagram <i>Class</i>	IV – 30
Gambar V–1. Kurva Hasil Pengujian Jumlah Klaster Optimum	V – 20

BAB I PENDAHULUAN

1.1 Pendahuluan

Bab ini menjelaskan secara garis besar sebab dilakukannya penelitian. Bab ini berisikan latar belakang masalah penelitian, rumusan masalah yang kemudian lebih dirincikan menjadi *research questions*, tujuan penelitian, manfaat penelitian, batasan masalah penelitian dan sistematika penulisan.

1.2 Latar Belakang Masalah

Clustering atau pengklasteran merupakan salah satu dari lima metode disiplin ilmu *data mining* yang merupakan bagian dari metode mayor. Pengklasteran adalah proses pengelompokan suatu himpunan objek data kedalam sejumlah klaster, sehingga objek-objek dalam satu klaster memiliki kesamaan (*similarity*) yang tinggi dan memiliki perbedaan (*dissimilarity*) yang tinggi dengan objek-objek pada klaster lain (Han, Kamber dan Pei, 2012). Menurut Santoso (2007) ada dua pendekatan dalam pengklasteran yaitu partisi dan hirarki.

Algoritma *k-means* adalah salah satu algoritma pengklasteran dengan pendekatan partisi yang dikembangkan oleh Mac Queen pada tahun 1967. Algoritma *k-means* merupakan metode analisis kelompok yang memisahkan data kedalam dua atau lebih kelompok secara iteratif dimana setiap objek dimiliki oleh sebuah kelompok dengan rata-rata (*mean*) terdekat (Prasetyo, 2012).

Algoritma k-menas populer digunakan dalam permasalahan pengklasteran karena sederhana dan mudah digunakan dalam praktiknya dan mampu menangani pengelompokan data berukuran besar dengan cepat serta efisien dalam penggunaan memori dan kompleksitasnya yang linier dengan jumlah data (Dong dan Qi, 2009; Saranya dan Krishnakumari, 2011; Alfina, Santosa dan Barakbah, 2012; P, Sunil dan Balachandran, 2013; Eka Merliana, Ernawati dan Santoso, 2015; Mishra, Saini dan Bagri, 2015; *et al.*, 2017).

Kualitas hasil pengklasteran algoritma *k-means* sangat tergantung pada inisialisasi *centroid* atau pusat kluster awal (P, Sunil dan Balachandran, 2013), inisialisasi pusat kluster awal yang tidak tepat dapat menyebabkan kualitas hasil pengklasteran menjadi kurang baik dan juga memiliki pengaruh pada jumlah iterasi dan waktu pemrosesan yang diperlukan saat menjalankan algoritma *k-means* (Saranya dan Krishnakumari, 2011). Pada umumnya inisialisasi pusat kluster awal algoritma *k-means* dibangkitkan secara acak, inisialisasi pusat kluster awal yang dibangkitkan secara acak sering kali menghasilkan kondisi konvergensi yang bersifat *local optimum* (Alfina, Santosa dan Barakbah, 2012). Selain itu, Inisialisasi pusat kluster awal yang dibangkitkan secara acak juga tidak menjanjikan menghasilkan pengelompokan yang konsisten (Saranya dan Krishnakumari, 2011).

Kondisi konvergensi yang bersifat *local optimum* pada algoritma *k-means* terjadi karena pengeksplorasi hanya terjadi pada sekitaran ruang yang sempit disekeliling pusat kluster (Cui, Potok dan Palathingal, 2005). Maka dari itu perlu dicarikan cara agar algoritma *k-means* dapat memenuhi kondisi konvergensi yang

bersifat *global optimum* dimana pengeplorasian terjadi untuk seluruh ruang pencarian (*search space*). Untuk memenuhi kondisi konvergensi yang bersifat *global optimum* tersebut dapat diperbantukan dengan metode-metode yang mempunyai sifat optimalisasi.

Menurut Suyanto (2010) ada dua pendekatan pada metode-metode optimalisasi yaitu deterministik dan probabilistik. Metode *Particle Swarm Optimization* (PSO) merupakan salah satu bagian dari metode optimalisasi dengan pendekatan probabilistik dari rumpun *swarm intelligence*. Metode PSO adalah metode optimasi berbasis populasi yang dikembangkan oleh James Kennedy dan Russ Eberhart pada tahun 1995. Kecerdasan PSO dalam menemukan kondisi yang bersifat *global optimum* didasari pada kecerdasan individu dan dipengaruhi oleh kecerdasan kawanan. Pada metode PSO jika satu partikel menemukan jalan yang optimal menuju sumber solusi, maka kawanan lainnya juga akan segera mengikuti jalan tersebut meskipun lokasi mereka berjauhan (Suyanto, 2010). Berdasarkan kehadalan metode PSO dalam menemukan solusi yang bersifat *global optimum* tersebut maka penelitian ini bertujuan untuk menguji pengaruh metode *Particle Swarm Optimization* dalam penentuan *centroid* awal terhadap kualitas hasil *clustering* algoritma *k-means*.

Selain itu, pada dasarnya kebanyakan algoritma pengklasteran akan baik dalam menangani data berdimensi rendah (Hidayati, Ihsan dan Danny, 2017). Karena pada data berdimensi tinggi akan terjadi fenomena *curse of dimensionality* dimana volume ruang meningkat cepat sehingga letak data menjadi jarang. Selain itu, pada pengklasteran dengan pendekatan partisi salah satu dampaknya adalah

meningkatnya kompleksitas dalam perhitungan jarak secara eksponensial. Selain itu, ketika dimensi tinggi biasanya hanya sebagian kecil dimensi yang relevan dengan kelompok tertentu, dimensi yang tidak relevan dapat menyebabkan kebisingan dan menutupi proses pengelompokan yang akan dibentuk (Saranya dan Krishnakumari, 2011).

Karena data yang digunakan pada penelitian ini berupa dokumen teks yang apabila dikonversi menjadi data numerik akan menghasilkan data yang memiliki dimensi tinggi (Kadhim, Cheah dan Ahamed, 2014), untuk itu sebelum dilakukan proses pengklasteran menggunakan algoritma *k-means* data tersebut terlebih dahulu dilakukan proses reduksi dimensi. Menurut penelitian Hidayati, Ihsan dan Danny (2017) dengan dilakukannya reduksi dimensi menggunakan metode *Singular Value Decomposition* (SVD) pada proses pengklasteran dapat meningkatkan kualitas hasil pengklasteran.

1.3 Rumusan Masalah

Berdasarkan latar belakang yang dikemukakan diatas, rumusan masalah pada penelitian ini adalah apa pengaruh metode *Particle Swarm Optimization* dalam penentuan *centroid* awal terhadap kualitas hasil *clustering* algoritma *k-means*.

Untuk memudahkan menjawab rumusan masalah yang disebutkan diatas, maka pada penelitian ini akan dibagi beberapa *research questions* antara lain:

1. Bagaimana mekanisme mengkonversi data dari teks menjadi numerik?
2. Bagaimana metode *Singular Value Decomposition* dapat melakukan reduksi dimensi?
3. Bagaimana metode *Particle Swarm Optimization* dapat menentukan pusat kluster awal algoritma *k-means*?
4. Bagaimana perbandingan kualitas hasil pengklasteran algoritma *k-means* dengan penentuan pusat kluster awal menggunakan metode *Particle Swarm Optimization* terhadap algoritma *k-means* dengan penentuan pusat kluster awal ditetapkan secara acak dengan data tanpa dilakukan reduksi dimensi?
5. Bagaimana perbandingan kualitas hasil pengklasteran algoritma *k-means* dengan penentuan pusat kluster awal menggunakan metode *Particle Swarm Optimization* terhadap algoritma *k-means* dengan penentuan pusat kluster awal ditetapkan secara acak dengan data dilakukan reduksi dimensi?

1.4 Tujuan Penelitian

Tujuan dilakukannya penelitian adalah sebagai berikut:

1. Mengetahui mekanisme konversi data dari teks menjadi numerik.
2. Mengetahui mekanisme metode *Singular Value Decomposition* dalam melakukan reduksi dimensi.
3. Mengetahui mekanisme metode *Particle Swarm Optimization* dalam menentukan pusat kluster awal algoritma *k-means*.
4. Mengetahui perbandingan kualitas hasil pengklasteran algoritma *k-means* dengan penentuan pusat kluster awal menggunakan metode *Particle Swarm*

Optimization terhadap algoritma *k-means* dengan penentuan pusat kluster awal ditetapkan secara acak dengan data tanpa dilakukan reduksi dimensi.

5. Mengetahui perbandingan kualitas hasil pengklasteran algoritma *k-means* dengan penentuan pusat kluster awal menggunakan metode *Particle Swarm Optimization* terhadap algoritma *k-means* dengan penentuan pusat kluster awal ditetapkan secara acak dengan data dilakukan reduksi dimensi.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut :

1. Memahami mekanisme konversi data dari teks menjadi numerik.
2. Memahami mekanisme metode *Singular Value Decomposition* dalam melakukan reduksi dimensi.
3. Memahami mekanisme metode *Particle Swarm Optimization* dalam penentuan pusat kluster awal algoritma *k-means*.
4. Memahami perbandingan kualitas hasil pengklasteran algoritma *k-means* dengan penentuan pusat kluster awal menggunakan metode *Particle Swarm Optimization* terhadap algoritma *k-means* dengan penentuan pusat kluster awal ditetapkan secara acak dengan data tanpa dilakukan reduksi dimensi.
5. Memahami perbandingan kualitas hasil pengklasteran algoritma *k-means* dengan penentuan pusat kluster awal menggunakan metode *Particle Swarm Optimization* terhadap algoritma *k-means* dengan penentuan pusat kluster awal ditetapkan secara acak dengan data dilakukan reduksi dimensi.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Data uji yang digunakan pada penelitian ini berupa dokumen jurnal berbahasa Indonesia.
2. Proses konversi data pada penelitian ini secara berurutan meliputi *case folding*, *remove punctuation*, *tokenizing*, *stemming*, *stop words removal* dan pembobotan yang menggunakan metode *Term Frequency And Inverse Document Frequency* (TF-IDF).
3. Proses Reduksi dimensi pada penelitian ini menggunakan metode *Singular Value Decomposition* (SVD).
4. Parameter acuan untuk perbandingan kualitas metode pengklasteran meliputi nilai *Davies Bouldin Index* (DBI), jumlah iterasi dan waktu komputasi.
5. Teknik perhitungan jarak pada metode *Particle Swarm Optimization* (PSO), algoritma *k-means* dan metode *Davies Bouldin Index* (DBI) menggunakan teknik *Euclidean Distance*.

1.7 Sistematika Penulisan

Sistematika penulisan penelitian ini adalah sebagai berikut :

BAB I. PENDAHULUAN

Pada bab ini akan diuraikan mengenai latar belakang masalah penelitian, rumusan masalah yang kemudian lebih dirincikan dalam *research*

questions, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini akan dibahas dasar-dasar teori dan penelitian yang relevan terhadap penelitian yang dilakukan.

BAB III. METODOLOGI PENELITIAN

Pada bab ini akan dibahas mengenai rencana dari tahapan-tahapan untuk menyelesaikan masalah pada penelitian. Setiap rencana akan dideskripsikan dengan rinci yang mengacu pada kerangka kerja. Pada akhir bab ini akan berisikan perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini akan dibahas mengenai tahapan-tahapan pengembangan perangkat lunak yang digunakan sebagai alat penelitian.

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini akan menyajikan hasil dari penelitian yang kemudian akan dilakukan analisis. Analisis ini nantinya digunakan sebagai basis dari kesimpulan yang diambil dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini berisikan kesimpulan dari penelitian yang dilakukan dan juga berisikan saran-saran yang diharapkan berguna pada penelitian selanjutnya.

1.8 Kesimpulan

Kualitas hasil pengklasteran algoritma *k-means* sangat tergantung pada inisialisasi pusat kluster awal (P, Sunil dan Balachandran, 2013), inisialisasi pusat kluster awal yang tidak tepat dapat menyebabkan kualitas hasil pengklasteran menjadi kurang baik dan memiliki pengaruh pada jumlah iterasi dan waktu pemrosesan yang diperlukan saat menjalankan algoritma *k-means* (Saranya dan Krishnakumari, 2011).

Pada umumnya inisialisasi pusat kluster awal pada algoritma *k-means* dibangkitkan secara acak, inisialisasi pusat kluster awal yang dibangkitkan secara acak sering kali menghasilkan kondisi konvergensi yang bersifat *local optimum* (Alfina, Santosa dan Barakbah, 2012). Maka dari itu perlu dicarikan cara agar algoritma *k-means* dapat memenuhi kondisi konvergensi yang bersifat *global optimum* dimana pengeksplorasian terjadi pada seluruh ruang pencarian (*search space*). Untuk memenuhi kondisi konvergensi yang bersifat *global optimum* dapat diperbantukan dengan metode-metode yang mempunyai sifat optimalisasi.

Particle Swarm Optimization (PSO) merupakan salah satu bagian dari metode optimalisasi dengan pendekatan probabilistik dari rumpun *swarm intelligence*. Kecerdasan PSO dalam menemukan kondisi yang bersifat *global optimum* didasari pada kecerdasan individu dan pengaruh dari kecerdasan kawanan (Suyanto, 2010). Berdasarkan kehadalan metode PSO dalam menemukan kondisi *global optimum* maka penelitian ini bertujuan untuk menguji pengaruh metode *Particle Swarm Optimization* dalam penentuan *centroid* awal terhadap kualitas hasil *clustering* algoritma *k-means*.

DAFTAR PUSTAKA

- 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.
- Cui, X., Potok, T. E. and Palathingal, P. (2005) 'Document Clustering using Particle Swarm Optimization', pp. 1–7.
- Dong, J. and Qi, M. (2009) 'K-means optimization algorithm for solving clustering problem', *Proceedings - 2009 2nd International Workshop on Knowledge Discovery and Data Mining, WKKD 2009*, pp. 52–55. doi: 10.1109/WKDD.2009.85.
- 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.
- Fabregas, A. C., Gerardo, B. D. and Tanguilig III, B. T. (2017) 'Enhanced Initial Centroids for K-means Algorithm', *International Journal of Information Technology and Computer Science*, 9(1), pp. 26–33. doi:

10.5815/ijitcs.2017.01.04.

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.

Kruchten, P. (2003) *Philippe Kruchten - The Rational Unified Process_ An Introduction-Addison-Wesley Professional (2003)*. third. Addison Wesley.

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.

Mu'arif, K. (2009) 'Komparasi pemodelan data menggunakan c4.5 dan c4.5 berbasis particle swarm optimization untuk memprediksi kelulusan mahasiswa'.

- 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. Available at: <http://www.ijser.org>.
- 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).
- Santika, P. W. *et al.* (2010) 'Implementasi Algoritma Particle Swarm Optimization untuk Penentuan Posisi Optimum Router-router Campus-wide WMN'.
- 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.

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. Available at: <http://journal.ilmukomputer.org>.