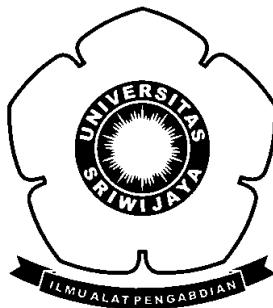


**PENGARUH ALGORITMA ARTIFICIAL BEE COLONY  
DALAM MENENTUKAN CENTROID AWAL  
ALGORITMA K-MEANS**

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



Oleh :

Nopren  
09021181520131

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

## LEMBAR PENGESAHAN TUGAS AKHIR

### PENGARUH ALGORITMA *ARTIFICIAL BEE COLONY* DALAM MENENTUKAN *CENTROID* AWAL ALGORITMA *K-MEANS*

Oleh :

NOPREN  
NIM : 09021181520131

Mengetahui,

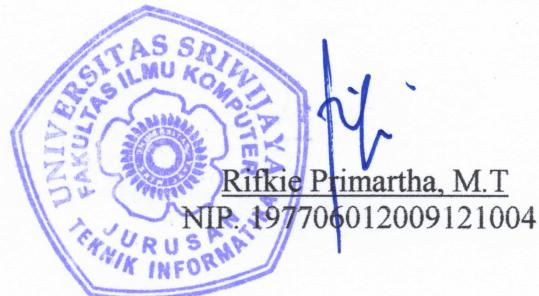
Indralaya, 12 Agustus 2020

Pembimbing I,



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

Mengetahui,  
Ketua Jurusan Teknik Informatika



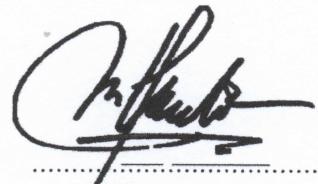
## TANDA LULUS SIDANG TUGAS AKHIR

Pada hari Kamis, 30 Juli 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Nopren  
NIM : 09021181520131  
Judul : Pengaruh Algoritma *Artificial Bee Colony* Dalam Menentukan *Centroid* Awal Algoritma *K-Means*

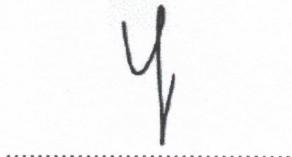
1. Pembimbing I

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



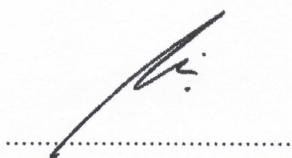
2. Penguji I

Yunita, M.Cs.  
NIP. 198306062015042002



3. Penguji II

Rizki Kurniati, M.T.  
NIPUS. 1671045207910003



Mengetahui,  
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T  
NIP. 197706012009121004

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Nopren

NIM : 09021181520131

Program Studi : Teknik Informatika

Judul Skripsi : Pengaruh Algoritma *Artificial Bee Colony* Dalam Menentukan  
*Centroid Awal Algoritma K-Means*

Hasil Pengecekan Software *iThenticate/Turnitin* : 2 %

Menyatakan bahwa Laporan Tugas Akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila 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



Nopren

NIM. 09021181520131



## **HALAMAN MOTTO DAN PERSEMPAHAN**

Motto :

**TERLIHAT BERANTAKKAN, NAMUN SEBENARNYA TERSUSUN RAPI !**

Kupersembahkan karya tulis ini kepada :

- Kedua orang tuaku
- Keluarga besar Dahuri
- Teman seperjuanganku
- Jurusan Teknik Informatika
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

# THE EFFECT OF ARTIFICIAL BEE COLONY ALGORITHM IN DETERMINE INITIAL CENTROID OF KMEANS ALGORITHM

by:  
Nopren  
09021181520131

## ABSTRACT

K-means clustering algorithm is an algorithm with the concepts of similarity and dissimilarity. In general, initial centroid determination of the k-means clustering algorithm is carried out randomly so that it allows optimum conditions to occur only to the nearest candidate solution (local optimum) not to all existing solutions (global optimum). Therefore, to improve the quality of clustering results, another algorithm is needed to help calculate the initial centroid value of the k-means algorithm. One optimization algorithm is the artificial bee colony algorithm, whose working process simulates the intelligent foraging behavior of bees. This study will compare the application of the Artificial Bee Colony algorithm to the initial centroid determination of the K-Means Clustering algorithm. The results obtained from the comparison of the quality of the results of grouping the k-means algorithm with the initial centroid optimized by the artificial bee colony algorithm experienced a percentage decrease in DBI value of 49.58% for data without reduction, and decreased by 58.15% for the data already reduced. On the other hand, the k-means algorithm with initial centroid optimized by the artificial bee colony algorithm is able to speed up the convergence process when compared to the k-means algorithm with random initial centroids.

**Keywords:** *Similarity, Dissimilarity, Centroid, Local Optimum, Global Optimum, K-Means Clustering, Artificial Bee Colony.*

# PENGARUH ALGORITMA ARTIFICIAL BEE COLONY DALAM MENENTUKAN CENTROID AWAL ALGORITMA K-MEANS

Oleh:  
Nopren  
09021181520131

## ABSTRAK

algoritma *k-means clustering* merupakan algoritma dengan konsep *similarity* dan *dissimilarity*. Pada umumnya penentuan *centroid* awal algoritma *k-means clustering* dilakukan secara acak sehingga memungkinkan terjadi kondisi optimum hanya pada kandidat solusi terdekatnya saja (*local optimum*) bukan pada semua solusi yang ada (*global optimum*). Oleh karena itu untuk meningkatkan kualitas hasil pengklasteran, dibutuhkan algoritma lain untuk membantu memperhitungkan nilai *centroid* awal algoritma *k-means*. Salah satu algoritma optimasi adalah algoritma *artificial bee colony*, yang proses kerjanya mensimulasikan perilaku mencari makan yang cerdas kawanan lebah. Penelitian ini akan membandingkan penerapan algoritma *Artificial Bee Colony* terhadap penentuan *centroid* awal algoritma *K-Means Clustering*. Hasil yang didapatkan dari perbandingan kualitas hasil pengklasteran algoritma *k-means* dengan *centroid* awal dioptimasi oleh algoritma *artificial bee colony* mengalami presentase penurunan nilai DBI sebesar 49,58% untuk data tanpa reduksi, dan mengalami penurunan sebesar 58.15% untuk data sudah direduksi. Disisi lain, algoritma *k-means* dengan *centroid* awal dioptimasi oleh algoritma *artificial bee colony* mampu mempercepat proses konvergen jika dibandingkan dengan algoritma *k-means* dengan *centroid* awal acak.

**Kata Kunci:** *Similarity, Dissimilarity, Centroid, Local Optimum, Global Optimum, K-Means Clustering, Artificial Bee Colony.*

## 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 dan lancar. 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 tua 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. Bapak M. Qurhanul Rizqie 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 penggerjaan Tugas Akhir.

6. Ibu Yunita, M.Cs. selaku penguji I dan Ibu Rizki Kurniati, M.T selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses penggerjaan Tugas Akhir.
7. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
8. Kak Ricy selaku admin jurusan teknik informatika 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

Nopren

## 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
ABSTRAKSI.....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR .....	xix
BAB I .....	I-1
1.1 Pendahuluan.....	I-1.
1.2 Latar Belakang Masalah .....	I-1.
1.3 Rumusan Masalah.....	I-6.
1.4 Tujuan Penelitian.....	I-7.
1.5 Manfaat Penelitian.....	I-8.
1.6 Batasan Masalah.....	I-8.
1.7 Sistematika Penulisan.....	I-9.
1.8 Kesimpulan.....	I-10.
BAB II .....	II-1
2.1 Pendahuluan.....	II-1.
2.2 Landasan Teori .....	II-1.

2.2.1 Konversi Data .....	II-2.
2.2.2 Algoritma <i>Artificial Bee Colony</i> .....	II-3.
2.2.3 Algoritma <i>K-means</i> .....	II-9.
2.2.4 Reduksi Dimensi.....	II-11.
2.2.5 <i>Singular Value Decomposition</i> (SVD) .....	II-12.
2.2.6 <i>Davies Bouldin Index</i> (DBI) .....	II-15.
2.2.7 <i>Rational Unified Process</i> (RUP).....	II-16.
2.3 Penelitian Lain yang Relevan .....	II-17.
2.4 Kesimpulan.....	II-19.
 BAB III.....	III-1
3.1 Pendahuluan.....	III-1.
3.2 Unit Penelitian dan Data.....	III-1.
3.3 Tahapan Penelitian .....	III-2.
3.3.1 Menentapkan Kerangka Kerja .....	III-2.
3.3.2 Menetapkan Kriteria Pengujian .....	III-11.
3.3.3 Menetapkan Format Data Pengujian.....	III-12.
3.3.3.1 Penetapan Jumlah <i>Cluster</i> Optimum.....	III-12.
3.3.3.2 Pengujian Parameter Algoritma <i>Artificial Bee Colony</i> dalam Menentukan Centroid Awal Algoritma <i>K-means</i> .....	III-13.
3.3.3.3 Pengujian Hasil Clustering Algoritma <i>K-means</i> dengan Penetapan Centroid Awal Secara Acak.....	III-14.
3.3.3.4 Pengujian Hasil Clustering Algoritma <i>K-means</i> dengan Penetapan Centroid Awal Menggunakan Algoritma <i>Artificial Bee Colony</i> .....	III-15.
3.3.4 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian.....	III-16.
3.3.5 Melakukan Pengujian Penelitian .....	III-17.
3.3.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan Penelitian .....	III-17.
3.4 Metode Pengembangan Perangkat Lunak .....	III-18.
3.4.1 <i>Rational Unified Process</i> (RUP).....	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-21.
3.5 Manajemen Proyek Penelitian .....	III-21.
3.6 Kesimpulan.....	III-28.
 BAB IV .....	IV-1
4.1 Pendahuluan.....	IV-1.
4.2 Fase Insepsi.....	IV-1.
4.2.1 Pemodelan Bisnis .....	IV-1.
4.2.2 <i>Requerements</i> .....	IV-5.
4.2.3 <i>Analysis and Design</i> .....	IV-15.
4.3 Fase Elaborasi .....	IV-17.
4.3.1 <i>Business Modeling</i> .....	IV-18.
4.3.2 <i>Requirements</i> .....	IV-30.
4.3.3 <i>Analysis and Design</i> .....	IV-29.
4.4 Fase Konstruksi .....	IV-35.
4.4.1 <i>Business Modeling</i> .....	IV-36.
4.4.2 <i>Requirements</i> .....	IV-38.
4.4.3 <i>Implementation</i> .....	IV-38.
4.4.4 Implementasi Kelas .....	IV-38.
4.5 Fase Transisi .....	IV-40.
4.5.1 <i>Business Modeling</i> .....	IV-40.
4.5.2 <i>Requirements</i> .....	IV-40.
4.5.3 <i>Testing</i> .....	IV-41.
4.5.3.1 Rencana Pengujian .....	IV-41.
4.5.3.2 Implementasi Pengujian .....	IV-44.
4.6 Kesimpulan .....	IV-53.
 BAB V .....	V-1
5.1 Pendahuluan.....	V-1.
5.2 Konfigurasi Percobaan .....	V-1.
5.3 Hasil Pengujian.....	V-3.
5.3.1 Hasil Pengujian <i>k</i> optimum.....	V-3.
5.3.2 Hasil Pengujian Pengklasteran K-Means dengan Centroid Awal Acak.	V-4.

5.3.3 Hasil Pengujian Parameter Algoritma <i>Artificial Bee Colony</i> .....	V-6.
5.3.3.1 Hasil Pengujian Jumlah Iterasi .....	V-6.
5.3.3.2 Hasil Pengujian Jumlah Koloni .....	V-10.
5.3.4 Hasil Pengujian Pengklasteran <i>K-Means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma <i>Artificial Bee Colony</i> .....	V-14.
5.4 Olah Data.....	V-15.
5.4.1 Olah Data Hasil Pengujian <i>k</i> Optimum .....	V-15.
5.4.2 Olah Data Hasil Perbandingan Nilai DBI.....	V-16.
5.4.3 Olah Data Hasil Perbandingan Waktu Komputasi .....	V-23.
5.4.4 Olah Data Hasil Perbandingan Jumlah Iterasi .....	V-29.
5.5 Analisis Hasil Pengujian.....	V-35.
5.5.1 Analisis Hasil Pengujian Nilai <i>k</i> optimum .....	V-35.
5.5.2 Analisis Hasil Pengujian Parameter Algoritma Artificial Bee Colony V	35.
5.5.3 Analisis Hasil Pengujian Pengklasteran <i>K-Means</i> dengan <i>Centroid</i> Awal Acak pada Data Hasil Reduksi dan Tanpa Reduksi.....	V-36.
5.5.4 Analisis Hasil Pengujian Pengklasteran <i>K-Means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma Artificial Bee Colony pada Data Hasil Reduksi dan Tanpa Reduksi .....	V-37.
5.5.5 Analisis Hasil Pengujian Pengklasteran <i>K-Means</i> dengan <i>Centroid</i> Awal Acak dan <i>Centroid</i> Awal dari Perhitungan Algoritma Artificial Bee Colony pada Data Reduksi Maupun Tanpa Reduksi Dimensi .....	V-38.
5.6 Kesimpulan.....	V-39.
BAB VI .....	VI-1
6.1 Pendahuluan.....	VI-1.
6.2 Kesimpulan.....	VI-1.
6.3 Saran .....	VI-2.
DAFTAR PUSTAKA .....	xxiii
LAMPIRAN 1 KODE PROGRAM .....	L-1

## DAFTAR TABEL

	Halaman
Tabel III – 1 Rancangan Tabel Penetapan Jumlah Klaster Optimum.....	III-15.
Tabel III – 2 Rancangan Tabel Hasil Pengujian Jumlah Iterasi Terbaik Dengan Jumlah Koloni Konstan.....	III-16.
Tabel III – 3 Rancangan Tabel Hasil Pengujian Jumlah Koloni Terbaik Dengan Jumlah Iterasi Konstan.....	III-16.
Tabel III – 4 Rancangan Tabel Hasil Pengujian <i>Clustering</i> Algoritma <i>K-Means</i> dengan <i>Centroid</i> Awal Secara Acak Sebelum Reduksi dan Sesudah Reduksi.....	III-17.
Tabel III– 5 Rancangan Tabel Hasil Pengujian <i>Clustering</i> Algoritma <i>K-Means</i> dengan <i>Centroid</i> Awal Menggunakan Algoritma <i>Artificial Bee Colony</i> Sebelum Reduksi dan Sesudah Reduksi .....	III-18.
Tabel III – 6 Penjadwalan penelitian dalam bentuk <i>Work Breakdown Structure</i>	III-24.
Tabel IV – 1 Kebutuhan Fungsional .....	IV-2.
Tabel IV – 2 Kebutuhan Non-Fungsional .....	IV-2.
Tabel IV – 3 Definisi Aktor .....	IV-3.
Tabel IV – 4 Definisi <i>Use Case</i> .....	IV-4.
Tabel IV – 5 Skenario <i>Use Case</i> Melakukan Konversi Data.....	IV-6.
Tabel IV – 6 Skenario <i>Use Case</i> Melakukan Pengklasteran Data Algoritma <i>K-</i> <i>means</i> dengan <i>Centroid</i> Awal Acak.....	IV-7.
Tabel IV – 7 Skenario <i>Use Case</i> Melakukan Pengklasteran Data Algoritma <i>K-</i> <i>means</i> dengan <i>Centroid</i> Awal Dari Perhitungan Algoritma <i>Artificial Bee Colony</i>	IV-9.

Tabel IV – 8 Skenario <i>Use Case</i> Melakukan Pengklasteran Data Reduksi Algoritma <i>K-means</i> dengan <i>Centroid</i> Awal Acak .....	IV-12.
Tabel IV – 9 Skenario <i>Use Case</i> Melakukan Pengklasteran Data Reduksi Algoritma <i>K-means</i> dengan <i>Centroid</i> Awal Dari Perhitungan Algoritma <i>Artificial Bee Colony</i> .....	IV-14.
Tabel IV – 10 Implementasi Kelas.....	IV-38.
Tabel IV – 11 Rencana Pengujian <i>Use Case</i> Melakukan Konversi Data .....	IV-41.
Tabel IV – 12 Rencana Pengujian <i>Use Case</i> Pengklasteran Data Algoritma <i>K-means</i> dengan <i>Centroid</i> Awal Acak.....	IV-42.
Tabel IV – 13 Rencana Pengujian <i>Use Case</i> Pengklasteran Data Algoritma <i>K-means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma <i>Artificial Bee Colony</i>	IV-43.
Tabel IV – 14 Rencana Pengujian <i>Use Case</i> Pengklasteran Data Reduksi Algoritma <i>K-means</i> dengan <i>Centroid</i> Awal Acak .....	IV-43.
Tabel IV – 15 Rencana Pengujian <i>Use Case</i> Pengklasteran Data Reduksi Algoritma <i>K-means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma <i>Artificial Bee Colony</i> .....	IV-44.
Tabel IV – 16 Pengujian Fitur Konversi Data .....	IV-45.
Tabel IV – 17 Rencana Pengujian <i>Use Case</i> Pengklasteran Data Algoritma <i>K-means</i> dengan <i>Centroid</i> Awal Acak.....	IV-46.
Tabel IV – 18 Pengujian Fitur Pengklasteran Data Algoritma <i>k-means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma <i>Artificial Bee Colony</i> .....	IV-47.
Tabel IV – 19 Pengujian Fitur Pengklasteran Data Reduksi Algoritma <i>k-means</i> dengan <i>Centroid</i> Awal Acak.....	IV-50.

Tabel IV – 20 Pengujian Fitur Pengklasteran Data Reduksi Algoritma <i>K-means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma <i>ArtificialBee Colony</i> .....	IV-52.
Tabel V – 1 Hasil Pengujian <i>k</i> Optimum .....	V-4.
Tabel V – 2 Hasil Pengujian Pengklasteran Data Algoritma <i>K-Means</i> Dengan <i>Centroid</i> Awal Acak Sebelum Dan Sesudah Reduksi .....	V-5.
Tabel V – 3 Hasil Pengujian Jumlah Iterasi Sebanyak 100 dan 200 Iterasi .....	V-7.
Tabel V – 4 Hasil Pengujian Jumlah Iterasi Sebanyak 300 dan 400 Iterasi .....	V-7.
Tabel V – 5 Hasil Pengujian Jumlah Iterasi Sebanyak 500 dan 600 Iterasi .....	V-8.
Tabel V – 6 Hasil Pengujian Jumlah Iterasi Sebanyak 700 dan 800 Iterasi .....	V-8.
Tabel V – 7 Hasil Pengujian Jumlah Iterasi Sebanyak 900 dan 1000 Iterasi .....	V-9.
Tabel V – 8 Hasil Pengujian Jumlah Koloni Sebanyak 10 dan 20 Koloni .....	V-11.
Tabel V – 9 Hasil Pengujian Jumlah Koloni Sebanyak 30 dan 40 Koloni .....	V-11.
Tabel V – 10 Hasil Pengujian Jumlah Koloni Sebanyak 50 dan 60 Koloni .....	V-12.
Tabel V – 11 Hasil Pengujian Jumlah Koloni Sebanyak 70 dan 80 Koloni .....	V-12.
Tabel V – 12 Hasil Pengujian Jumlah Koloni Sebanyak 90 dan 100 Koloni .....	V-13.
Tabel V – 13 Hasil Pengujian Pengklasteran Data Algoritma <i>K-Means</i> Dengan <i>Centroid</i> Awal Acak Sebelum Dan Sesudah Reduksi .....	V-14.
Tabel V – 14 Deskripsi Nilai DBI.....	V-16.
Tabel V – 15 Hasil Uji Normalitas Nilai DBI.....	V-17.
Tabel V – 16 Hasil Uji <i>Kruskal-Wallis</i> Perbandingan Nilai DBI .....	V-18.
Tabel V – 17 Hasil Uji <i>Mann-Whitney</i> Perbandingan Nilai DBI pada Data Hasil Pengklasteran K-Means dan K-Means+SVD .....	V-19.

Tabel V – 18 Hasil Uji <i>Mann-Whitney</i> Perbandingan Nilai DBI pada Data Hasil Pengklasteran <i>K-Means</i> dan <i>K-Means+ABC</i> .....	V-19.
Tabel V – 19 Hasil Uji <i>Mann-Whitney</i> Perbandingan Nilai DBI pada Data Hasil Pengklasteran K-Means dan K-Means+ABC+SVD .....	V-20.
Tabel V – 20 Hasil Uji <i>Mann-Whitney</i> Perbandingan Nilai DBI pada Data Hasil Pengklasteran K-Means+SVD dan K-Means+ABC .....	V-21.
Tabel V – 21 Hasil Uji <i>Mann-Whitney</i> Perbandingan Nilai DBI pada Data Hasil Pengklasteran K-Means+SVD dan K-Means+ABC+SVD .....	V-21.
Tabel V – 22 Hasil Uji <i>Mann-Whitney</i> Perbandingan Nilai DBI pada Data Hasil Pengklasteran K-Means+ABC dan K-Means+ABC+SVD.....	V-22.
Tabel V – 23 Deskripsi Waktu Komputasi .....	V-23.
Tabel V – 24 Hasil Uji Normalitas Waktu Komputasi .....	V-24.
Tabel V – 25 Hasil Uji <i>Kruskal-Wallis</i> Perbandingan Waktu Komputasi.....	V-25.
Tabel V – 26 Hasil Uji <i>Mann-Whitney</i> Perbandingan Waktu Komputasi pada Data Hasil Pengklasteran K-Means dan K-Means+SVD .....	V-25.
Tabel V – 27 Hasil Uji <i>Mann-Whitney</i> Perbandingan Waktu Komputasi pada Data Hasil Pengklasteran K-Means dan K-Means+ABC .....	V-26.
Tabel V – 28 Hasil Uji <i>Mann-Whitney</i> Perbandingan Waktu Komputasi pada Data Hasil Pengklasteran K-Means dan K-Means+ABC+SVD .....	V-26.
Tabel V – 29 Hasil Uji <i>Mann-Whitney</i> Perbandingan Waktu Komputasi pada Data Hasil Pengklasteran K-Means+SVD dan K-Means+ABC .....	V-27.
Tabel V – 30 Hasil Uji <i>Mann-Whitney</i> Perbandingan Waktu Komputasi pada Data Hasil Pengklasteran K-Means+SVD dan K-Means+ABC+SVD.....	V-28.

Tabel V – 31 Hasil Uji <i>Mann-Whitney</i> Perbandingan Waktu Komputasi pada Data Hasil Pengklasteran K-Means+ABC dan K-Means+ABC+SVD.....	V-28.
Tabel V – 32 Deskripsi Jumlah Iterasi.....	V-29.
Tabel V – 33 Hasil Uji Normalitas Jumlah Iterasi .....	V-30.
Tabel V – 34 Hasil Uji <i>Kruskal-Wallis</i> Perbandingan Jumlah Iterasi .....	V-30.
Tabel V – 35 Hasil Uji <i>Mann-Whitney</i> Perbandingan Jumlah Iterasi pada Data Hasil Pengklasteran K-Means dan K-Means+SVD .....	V-31.
Tabel V – 36 Hasil Uji <i>Mann-Whitney</i> Perbandingan Jumlah Iterasi pada Data Hasil Pengklasteran K-Means dan K-Means+ABC .....	V-32.
Tabel V – 37 Hasil Uji <i>Mann-Whitney</i> Perbandingan Jumlah Iterasi pada Data Hasil Pengklasteran K-Means dan K-Means+ABC+SVD.....	V-32.
Tabel V – 38 Tabel Hasil Uji <i>Mann-Whitney</i> Perbandingan Jumlah Iterasi pada Data Hasil Pengklasteran K-Means+SVD dan K-Means+ABC .....	V-33.
Tabel V – 39 Tabel Hasil Uji <i>Mann-Whitney</i> Perbandingan Jumlah Iterasi pada Data Hasil Pengklasteran K-Means+SVD dan K-Means+ABC+SVD.....	V-33.
Tabel V – 40 Tabel Hasil Uji <i>Mann-Whitney</i> Perbandingan Jumlah Iterasi pada Data Hasil Pengklasteran K-Means+ABC dan K-Means+ABC+SVD.....	V-34.

## DAFTAR GAMBAR

Halaman

Gambar II – 1 Ilustratif mekanisme optimasi algoritma <i>artificial bee colony</i> .....	II – 5.
Gambar II – 2 <i>Flowchart</i> inisialisasi <i>centroid</i> algoritma <i>Artificial Bee Colony</i> ....	II – 9.
Gambar II – 3 <i>Flowchart</i> Algoritma <i>K-means</i> .....	II – 11.
Gambar II – 4 Representasi <i>Singular Value Decomposition</i> (SVD).....	II – 13.
Gambar II – 5 Arsitektur <i>Rational Unified Process (RUP)</i> .....	II – 17.
Gambar III – 1. Tahapan Penelitian .....	III – 2.
Gambar III – 2. Kerangka Kerja Penelitian .....	III – 4.
Gambar III – 3. Penjadwalan Penelitian Tahap Menentukan Ruang Lingkup dan Unit Penelitian.....	III – 28.
Gambar III – 4. Panjadwalan Penelitian Tahap Menentukan Dasar Teori Berkaitan dengan Penelitian .....	III – 28.
Gambar III – 5. Penjadwalan Penelitian Tahap Menentukan Metodologi Pengembangan Perangkat Lunak Fase <i>Inception</i> .....	III – 29.
Gambar III – 6. Penjadwalan Penelitian Tahap Menentukan Metode Pengembangan Perangkat Lunak Fase <i>Elaboration</i> .....	III – 29.
Gambar III – 7. Penjadwalan Penelitian Tahap Menentukan Metode Pengembangan Perangkat Lunak Fase <i>Construction</i> .....	III – 29.
Gambar III – 8. Penjadwalan Penelitian Tahap Menentukan Metode Pengembangan Perangkat Lunak Fase <i>Transition</i> .....	III – 30.
Gambar III – 9. Penjadwalan Penelitian Tahap Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan .....	III – 30.

Gambar IV – 1 Diagram <i>Use Case</i> .....	IV – 3.
Gambar IV – 2 Diagram <i>Sequence</i> Konversi Data .....	IV – 19.
Gambar IV – 3 Diagram <i>Sequence</i> Pengklasteran Data Algoritma <i>K-means</i> dengan <i>Centroid</i> Awal Acak.....	IV – 20.
Gambar IV – 4 Diagram <i>Sequence</i> Pengklasteran Data Algoritma <i>K-means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma <i>Artificial Bee Colony</i> .....	IV – 21.
Gambar IV – 5 Diagram <i>Sequence</i> Pengklasteran Data Reduksi Algoritma <i>K-means</i> dengan <i>Centroid</i> Awal Acak.....	IV – 22.
Gambar IV – 6 Diagram <i>Sequence</i> Pengklasteran Data Reduksi Algoritma <i>K-means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma <i>Artificial Bee Colony</i>	IV – 23.
Gambar IV – 7 Fitur-fitur Perangkat Lunak .....	IV – 24.
Gambar IV – 8 Fitur Memasukkan Data.....	IV – 25.
Gambar IV – 9 Fitur Menampilkan Jumlah Dokumen dan <i>Term</i> .....	IV – 25.
Gambar IV – 10 Fitur Memilih Metode Pengelompokan .....	IV – 26.
Gambar IV – 11 Fitur Memasukkan Parameter Algoritma <i>K-means</i> .....	IV – 27.
Gambar IV – 12 Fitur Memasukkan Parameter Metode ABC.....	IV – 28.
Gambar IV – 13 Fitur Menjalankan Proses Pengklasteran .....	IV – 28.
Gambar IV – 14 Bagian Menampilkan <i>Centroid</i> Akhir.....	IV – 29.
Gambar IV – 15 Bagian Menampilkan Hasil Pengklasteran .....	IV – 29.
Gambar IV – 16 Bagian Menampilkan Hasil Perhitungan Nilai DBI, Waktu Komputasi dan Jumlah Iterasi.....	IV – 30.
Gambar IV – 17 Diagram Aktivitas Konversi Data.....	IV – 31.

Gambar IV – 18 Diagram Aktivitas Pengklasteran Data Algoritma <i>K-means</i> dengan <i>Centroid</i> Awal Acak.....	IV – 32.
Gambar IV – 19 Diagram Aktivitas Pengklasteran Data Algoritma <i>K-means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma <i>Artificial Bee Colony</i> .....	IV – 33.
Gambar IV – 20 Diagram Aktivitas Pengklasteran Data Reduksi Algoritma <i>K-means</i> dengan <i>Centroid</i> Awal Acak.....	IV – 34.
Gambar IV – 21 Diagram Aktivitas Pengklasteran Data Reduksi Algoritma <i>K-means</i> dengan <i>Centroid</i> Awal dari Perhitungan Algoritma <i>Artificial Bee Colony</i> .....	IV – 35.
Gambar IV – 22 Diagram Kelas.....	IV – 37.
Gambar IV – 1 Kurva Hasil Penentuan $k$ Optimum .....	IV – 16.

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Pendahuluan**

Pada BAB I akan menjelaskan secara umum penelitian yang akan dilakukan. Penjelasan berupa pembahasan latar belakang masalah, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan dan kesimpulan. Secara umum BAB I ini menjelaskan mengenai pengertian data mining, pengklasteran, dan pembahasan metode *k-means* yang *centroid* awalnya dibangkitkan secara acak akan dibantu oleh perhitungan algoritma *artificial bee colony*, yang terangkum dalam subbab 1.2.

#### **1.2 Latar Belakang Masalah**

Data mining adalah proses *learning* pada data yang dirancang untuk mencari suatu pengetahuan tersembunyi yang berharga dan konsisten (Gorunescu, 2011). Data mining, sering juga disebut sebagai *knowledge discovery in database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santosa, 2007).

Menurut Susanto and Suryadi, (2010), Data Mining dibagi menjadi dua fungsi, yaitu fungsi utama (*major*) dan fungsi tambahan (*minor*). *Klasifikasi*, pengelompokan atau pengklasteran dan *asosiasi* masuk ke kelompok fungsi *major* data mining, Sedangkan *deskripsi*, *estimasi* dan *prediksi* masuk ke fungsi *minor* data mining.

Fungsi *major* dalam data mining salah satunya adalah pengklasteran atau pengelompokan. Menurut Karaboga and Ozturk, (2010), tujuan pengklasteran adalah untuk mengelompokkan data ke dalam klaster sehingga kesamaan antara anggota data dalam klaster yang sama adalah maksimal sementara kesamaan di antara anggota data dari klaster yang berbeda sangat minim. Pengklasteran merupakan sebuah metode pengelompokan suatu objek kedalam sejumlah kelompok (klaster) yang sesuai. Pengklasteran dapat digunakan untuk mengidentifikasi kelompok-kelompok yang sebelumnya tidak terlihat dari data yang dihasilkan oleh algoritma pengklasteran. Karena alasan ini, algoritma pengklasteran telah mendapatkan popularitas di berbagai sektor seperti kesehatan, pendidikan, biologi, bisnis, keuangan, pemasaran, telekomunikasi, pemrosesan gambar (Ahmed and Rahim, 2017).

Berdasarkan jenisnya metode pengklasteran dibagi menjadi dua jenis yaitu jenis *hirarki* dan *non-hirarki* (Suyanto, 2010). Beberapa metode pengklasteran *hirarki* yang popular adalah C4.5, dan metode pengklasteran *non-hirarki* yang populer *k-means*, *k-medoids*, *Fuzzy C-means*.

Salah satu metode pengklasteran *non-hirarki* yang popular adalah algoritma *k-means*. Algoritma *k-means* ini merupakan salah satu algoritma klasterisasi yang

umum digunakan karena mudah untuk di adaptasi, diimplementasikan dan di jalankan serta waktu yang dibutuhkan untuk menjalankan pembelajaran ini relatif cepat. Algoritma *k-means* termasuk dalam *unsupervised learning* yang menerima masukan berupa data atau objek kemudian mengelompokan data atau objek tersebut ke dalam  $k$  buah kelompok. Pada setiap klaster terdapat *centroid* yang merepresentasikan klaster tersebut.

Penentuan *centroid* awal algoritma *k-means* pada umumnya sering kali dilakukan secara acak. Menurut Ahmed and Rahim, (2017), hasil dari algoritma *k-means* sangat tergantung pada estimasi *centroid* awal, yang berarti algoritma *k-means* standard mungkin memerlukan lebih banyak iterasi sampai posisi *centroid* stabil, jika *centroid* awal tidak dipilih secara efektif maka kualitas pengelompokan dapat menurun. Permasalahannya hasil klaster ini sering mengacu pada solusi *local optimum* (Nugroho, Purwitasari and Faticahah, 2016), yaitu solusi yang optimal (baik maksimal maupun minimal) pada kandidat solusi di tetangga terdekatnya saja bukan keseluruhan dari semua solusi yang ada atau yang biasa disebut *global optimum*.

Dalam buku karangan Han, Kamber and Pei, (2012), dijelaskan bahwa konsep pengklasteran adalah mengelompokan objek-objek kedalam suatu kelompok memiliki kesamaan yang tinggi (*similarity*) dan memiliki perbedaan yang tinggi dengan objek-objek pada kelompok lain (*dissimilarity*). Jadi pada kondisi *local optimum* hanya menghitung nilai pada objek-objek dalam satu klaster saja (*similarity*), sedangkan *global optimum* akan menghitung nilai objek-

objek dalam satu klaster (*similarity*) serta objek-objek pada klaster lainnya juga (*dissimilarity*).

Kondisi *local optimum* akan mempengaruhi hasil pengklasteran algoritma *k-means* menjadi kurang baik. Untuk mengatasi hal tersebut diperlukan kondisi yang *global optimum*, dimana untuk mendapatkan kondisi *global optimum* tersebut dapat dibantu dengan metode-metode yang mempunyai sifat optimalisasi. Menurut Suyanto, (2010), ada dua pendekatan optimalisasi berdasarkan metode operasinya yaitu *deterministic* dan *probabilistic*.

Metode optimasi dengan pendekatan *probabilistic* diantaranya adalah algoritma *particle swarm optimization* (PSO), algoritma *artificial bee colony* (ABC), algoritma *ant colony optimization* (ACO) dll. Metode optimasi yang dipakai pada penelitian ini adalah algoritma *artificial bee colony* (ABC) karena algoritma ini sederhana dalam konsep, mudah diimplementasikan, dan memiliki lebih sedikit parameter (Zou *et al.*, 2010).

Algoritma *artificial bee colony* adalah salah satu algoritma yang baru diperkenalkan, algoritma ini mensimulasikan berdasarkan perilaku mencari makan yang cerdas dari kawanan lebah madu. Tujuan lebah dalam model *artificial bee colony* adalah untuk menemukan solusi terbaik, posisi sumber makanan mewakili kemungkinan solusi untuk masalah optimasi dan nektar jumlah sumber makanan sesuai dengan kualitas (*fitness*) dari solusi terkait (Karaboga, Okdem and Ozturk, 2012). Menurut Zou *et al.* (2010), *Artificial bee colony* memiliki tiga kategori lebah yaitu *employed bee*, *onlooker bee* dan *scout bee*.

*Employed bee* adalah lebah yang berhubungan dengan sumber makanan tertentu yang sedang mereka eksplorasi, kemudian *employed bee* akan memberikan informasi tentang sumber makanan tersebut kepada *onlooker bee*. *Onlooker bee* adalah lebah yang menunggu di area dansa di sarang untuk menerima informasi yang diberikan oleh *employed bee* tentang sumber makanan. *Employed bee* yang memiliki sumber makanan namun telah di tinggalkan oleh kawanan lebah akan menjadi *scout bee* yang akan memulai pencarian acak untuk menemukan sumber makanan baru.

Menurut Nugroho, Purwitasari and Faticahah, (2016), *Onlooker bee* akan membandingkan nilai probabilitas masing-masing letak sumber makanannya dan mencari sumber makanan disekitar sumber yang dipilih, apabila sumber makanan yang ditemukan saat ini lebih banyak dari sumber sebelumnya maka lebah tersebut akan melupakan informasi tentang sumber makanan terbanyak sebelumnya.

Untuk sumber makanan yang habis atau tidak berubah, sumber makanan tersebut akan ditinggalkan dan dilupakan serta lebah tersebut akan menjadi *scout bee* yang akan mengintai sebuah sumber makanan baru yang akan diciptakan di ruang pencarian. Langkah-langkah tersebut akan dilakukan hingga ditemukan solusi *global optima*.

Dengan kemampuan yang dimiliki algoritma *Artificial bee colony*, yaitu dengan sifat pencarian secara global, maka penelitian ini di lakukan untuk menguji pengaruh algoritma *artificial bee colony* dalam menentukan *centroid* awal algoritma *k-means*.

Menurut Hidayati, Jambak and Saputra, (2017), pada dasarnya algoritma pengklasteran baik dalam menangani data berdimensi rendah. 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. Selain itu, algoritma pengklasteran sering terjebak dalam *curse of dimensionality* pada data berdimensi tinggi, dimana kompleksitas perhitungan jarak akan meningkat secara eksponensial sebanyak dimensi data (Saranya and Krishnakumari, 2011).

Penelitian ini akan menggunakan data dokumen teks jurnal berbahasa Indonesia. Menurut Kadhim, Cheah and Ahamed, (2014), apabila dokumen teks dikonversi menjadi data numerik maka akan menghasilkan data yang memiliki dimensi yang tinggi. Maka dari itu data tersebut akan dilakukan proses reduksi dimensi terlebih dahulu sebelum dilakukan proses pengklasteran. Dengan melakukan reduksi dimensi metode-metode pengklasteran dengan menggunakan metode *Singular Value Decomposition* (SVD), mampu meningkatkan kinerja pengklasteran *k-means* (Hidayati, Jambak and Saputra, 2017).

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang dikemukakan diatas, rumusan masalah dalam penelitian ini adalah bagaimana pengaruh *artificial bee colony* dalam menentukan *centroid* awal algoritma *k-means*.

Untuk menyelesaikan masalah di atas maka disusun menjadi dua pertanyaan penelitian (Research Question) :

1. Bagaimana algoritma *artificial bee colony* dapat menentukan *centroid* awal algoritma *k-means* ?
2. Bagaimana perbandingan kualitas hasil pengklasteran algoritma *k-means* dengan penentuan *centroid* awal secara acak dan penentuan *centroid* awal menggunakan algoritma *artificial bee colony* sebelum direduksi ?
3. Bagaimana perbandingan kualitas hasil pengklasteran algoritma *k-means* dengan penentuan *centroid* awal secara acak dan penentuan *centroid* awal menggunakan algoritma *artificial bee colony* sesudah direduksi ?

### **1.3 Tujuan Penelitian**

Tujuan dari penelitian ini adalah :

1. Mengetahui cara kerja algoritma *artificial bee colony* dalam menentukan *centroid* awal algoritma *k-means*
2. Mengetahui perbandingan kualitas hasil pengklasteran algoritma *k-means* dengan penentuan *centroid* awal secara acak dan penentuan *centroid* awal menggunakan algoritma *artificial bee colony* sebelum direduksi.
3. Mengetahui perbandingan kualitas hasil pengklasteran algoritma *k-means* dengan penentuan *centroid* awal secara acak dan penentuan *centroid* awal menggunakan algoritma *artificial bee colony* sesudah direduksi.

## 1.4 Manfaat Penelitian

Manfaat penelitian ini adalah sebagai berikut :

1. Memahami mekanisme kerja algoritma *artificial bee colony* dalam menentukan *centroid* awal algoritma *k-means*.
2. Mengetahui perbandingan hasil pengklasteran dengan *centroid* awal yang acak dan *centroid* awal yang ditentukan oleh algoritma *artificial bee colony*.
3. Memahami cara menguji hasil penggabungan algoritma *artificial bee colony* dengan metode *k-means* dalam menentukan *centroid* awal.
4. Memperoleh mekanisme baru pengoptimisasian algoritma *k-means* dengan menggunakan algoritma *artificial bee colony*.

## 1.5 Batasan Masalah

Dalam penelitian ini ada beberapa batasan masalah, yaitu:

1. Data pada penelitian ini menggunakan data dokumen teks berbahasa Indonesia.
2. Fokus penelitian adalah penentuan *centroid* awal algoritma *k-means*.
3. Algoritma yang digunakan untuk menentukan *centroid* awal algoritma *k-means* adalah algoritma *artificial bee colony*.
4. Reduksi dimensi menggunakan metode *Singular Value Decomposition* (SVD).

5. Perhitungan jarak algoritma *k-means* dan *artificial bee colony* menggunakan teknik *euclidian distance*.
6. Parameter yang digunakan untuk perbandingan kualitas pengklasteran adalah nilai DBI, jumlah iterasi dan waktu komputasi.

## **1.6 Sistematika Penulisan**

Sistematika penulisan penelitian ini adalah sebagai berikut :

### **BAB I. PENDAHULUAN**

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah/ruang lingkup, metodologi penelitian, dan sistematika penulisan.

### **BAB II. KAJIAN LITERATUR**

Pada bab ini akan dibahas dasar-dasar teori yang digunakan dalam penelitian, seperti definisi-definisi metode *k-means*, algoritma *artificial bee colony* dan *Davies-Bouldin Index*.

### **BAB III. METODOLOGI PENELITIAN**

Pada bab ini akan dibahas mengenai 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

Pada bab ini akan dibahas mengenai perancangan dan lingkungan implementasi penggabungan algoritma *artificial bee colony* dan metode *k-means*, hasil eksekusi, dan hasil pengujian.

## BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini, hasil pengujian berdasarkan langkah-langkah yang telah direncanakan disajikan. Analisis diberikan sebagai basis dari kesimpulan yang diambil dalam penelitian ini.

## BAB VI. KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan dari semua uraian-uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan berguna dalam penelitian selanjutnya.

### 1.7 Kesimpulan

Penentuan *centroid* awal algoritma *k-means* standard seringkali dilakukan secara acak, sehingga memungkinkan menghasilkan kondisi konvergensi yang baik tetapi hanya untuk klaster itu sendiri, bukan dari seluruh solusi yang ada atau yang biasa disebut *local optimum*. Sedangkan dikatakan pengklasteran yang baik adalah (*similarity* dan *disimilarity*) yaitu mirip dengan tetangga dekatnya yang disebut satu klaster dan tidak mirip dengan klaster yang lain. Dengan demikian diperlukan algoritma lain yang mampu mengoptimasi algoritma *k-means* agar

tidak mengalami kondisi *local optimum* tetapi mampu mencapai kondisi *global optimum*.

Salah satu metode optimasi dengan pendekatan *probabilistic* adalah algoritma *artificial bee colony* yang memiliki kemampuan untuk pencarian global dengan mensimulasikan mencari makanan yang cerdas dari kawanan lebah madu. Tujuan algoritma *artificial bee colony* adalah untuk menemukan solusi terbaik dari keseluruhan solusi yang ada. Dengan kemampuan yang dimiliki algoritma *artificial bee colony*, yaitu dengan sifat pencarian secara global, maka penelitian ini di lakukan untuk menguji pengaruh yang diberikan algoritma *artificial bee colony* dalam menentukan titik *centroid* awal 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’.
- 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.
- Baswade, A. M. and Nalwade, P. S. (2013) ‘Selection of Initial Centroids for k-Means Algorithm’, 2(July), pp. 161–164.
- Basyarah, A. N. *et al.* (2016) ‘INISIALISASI PUSAT CLUSTER MENGGUNAKAN ARTIFICIAL BEE COLONY PADA ALGORITMA POSSIBILISTIC FUZZY C-MEANS UNTUK SEGMENTASI CITRA’, 6, pp. 118–127.
- Cui, X., Potok, T. E. and Palathingal, P. (2005) ‘Document Clustering using Particle Swarm Optimization’, pp. 1–7.
- Dash, Rajashree and Dash, Rasmita (2012) ‘COMPARATIVE ANALYSIS OF K-MEANS AND GENETIC ALGORITHM BASED DATA CLUSTERING’, 3(June 2012), pp. 257–265.
- 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.
- Gorunescu, F. (2011) *Data Mining Concepts, models and Techniques*. volume 12. Berlin: Springer-Verlag. doi: 10.1007/978-3-642-19721-5.
- Han, J., Kamber, M. and Pei, J. (2012) *DATA MINING 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., Jambak, M. I. and Saputra, D. 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.
- Karaboga, D., Okdem, S. and Ozturk, C. (2012) ‘Cluster based wireless sensor

- network routing using artificial bee colony algorithm', pp. 847–860. doi: 10.1007/s11276-012-0438-z.
- Karaboga, D. and Ozturk, C. (2010) 'Fuzzy clustering with artificial bee colony algorithm', 5(14), pp. 1899–1902.
- Karaboga, D. and Ozturk, C. (2011) 'A novel clustering approach: Artificial Bee Colony (ABC) algorithm', *Applied Soft Computing Journal*, 11(1), pp. 652–657. doi: 10.1016/j.asoc.2009.12.025.
- 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.
- Nugroho, A. B., Purwitasari, D. and Faticahah, C. (2016) 'Implementasi Artificial Bee Colony untuk Pemilihan Titik Pusat pada Algoritma K-Means', *Jurnal Teknik ITS*, 5(2), p. A608.
- 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.
- Premkumar, M. S. and Ganesh, S. H. (2017) 'A Median based External Initial Centroid Selection Method for K-means Clustering', pp. 1–4. doi: 10.1109/WCCCT.2016.42.
- Raval, U. R. and Jani, C. (2016) 'Implementing & Improvisation of K-means Clustering Algorithm', *International Journal of Computer Science and Mobile Computing*, 55(5), pp. 191–203.
- Rendón, E. et al. (2011) 'Internal versus External cluster validation indexes', 5(1).
- Santosa, B. (2007) *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Pertama. Yogyakarta: Graha Ilmu.
- Santosa, B. and Umam, A. (2018) *Data Mining dan Big Data Analytics*. pertama. Edited by Isa. Penebar Media Pustaka.
- Saranya, M. and Krishnakumari, P. (2011) 'An Efficient Algorithm to fix Initial Centroid for Clustering High Dimensional data', 2(4), pp. 595–598.
- Susanto, S. and Suryadi, D. (2010) *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*. Edited by Nikodemus WK. Yogyakarta.
- Suyanto (2010a) *ALGORITMA OPTIMASI (Deterministik atau Probabilistik)*. pertama. graha ilmu.
- Suyanto (2010b) *Algoritma Optimasi Deterministik atau Probabilitik*. pertama.

Yogyakarta: Graha Ilmu.

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.

Zou, W. *et al.* (2010) ‘A Clustering Approach Using Cooperative Artificial Bee Colony Algorithm’, *Discrete Dynamics in Nature and Society*, 2010, pp. 1–17. doi: 10.1155/2010/459796.