

PERBANDINGAN ALGORITMA *CLUSTERING KERNEL K-MEANS* DAN *SPECTRAL CLUSTERING* DALAM KASUS *NON-LINEAR SEPARABLE CLUSTERING*

Diajukan Sebagai Syarat Untuk Menyelesaikan  
Pendidikan Program Strata-1 Pada  
Jurusan Teknik Informatika



Oleh :

Ilham Eldi Aldino  
09021381520061

**Jurusan Teknik Informatika**  
**FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA**  
**2020**

**LEMBAR PENGESAHAN TUGAS AKHIR**

**PERBANDINGAN ALGORITMA *CLUSTERING KERNEL K-MEANS* DAN  
*SPECTRAL CLUSTERING* DALAM KASUS *NON-LINEAR SEPARABLE*  
*CLUSTERING***

Oleh :

**ILHAM ELDI ALDINO  
NIM : 09021381520061**

Palembang, Juli 2020

Pembimbing I,

Pembimbing II



Rifkie Primartha, S.T., M.T  
NIP.197706012009121004

Danny Matthew Saputra, M.Sc.  
NIP.198505102015041002

Mengetahui,  
Ketua Jurusan Teknik Informatika,



Rifkie Primartha, M.T  
NIP. 197706012009121004

## TANDA LULUS UJIAN SIDANG TUGAS AKHIR

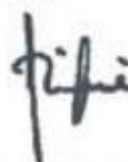
Pada hari Jumat tanggal 24 Juli 2020 telah dilaksanakan ujian sidang tugas akhir oleh jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Ilham Eldi Aldino

NIM : 09021381520061

Judul : Perbandingan Algoritma *Clustering Kernel K-Means* dan *Spectral Clustering* Dalam Kasus *Non-linear Separable Clustering*

1. Pembimbing I



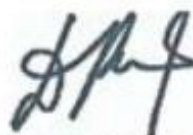
Rifkie Primartha, S.T., M.T  
NIP.197706012009121004

2. Pembimbing II



Danny Matthew Saputra, M.Sc  
NIP.198505102015041002

3. Penguji I



Dian Palupi Rini, M.Kom., Ph.D  
NIP. 197706012009121004

4. Penguji II



Kanda Januar Miraswaran, M.T  
NIPUS. 1671080901900006

Mengetahui,

Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T  
NIP. 197706012009121004

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Ilham Eldi Aldino  
NIM : 09021381520061  
Program Studi : Teknik Informatika  
Judul Skripsi : Perbandingan Algoritma *Clustering Kernel K-Means* dan *Spectral Clustering* Dalam Kasus *Non-linear Separable Clustering*  
Hasil Pergecekan Software *iThenticate/Turnitin* : 10 %

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

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



Palembang, Juli 2020



Ilham Eldi Aldino  
NIM. 09021381520061

*Motto :*

- *Every single thing has it's own price*

*Kupersembahkan karya tulis ini kepada :*

- *Orang tuaku dan kedua saudari kecilku yang senantiasa mendukungku*
- *Sahabat beserta teman-temanku yang setia untuk saling membantu*
- *Fakultas Ilmu Komputer*
- *Universitas Sriwijaya*

# COMPARISON OF CLUSTERING ALGORITHM BETWEEN KERNEL K-MEANS AND SPECTRAL CLUSTERING IN CASE OF NON-LINEAR SEPARABLE CLUSTERING

by :

Ilham Eldi Aldino

09021381520061

## ABSTRACT

Cluster analysis is a process that groups objects, this process produces a group consisting of many objects which can also be called as cluster, which inside the cluster there can be found similar objects which have the same characteristics as the others and have a different characteristic to a different cluster. Regarding the cluster data distribution, the problematics can be divided into two parts, it's called as linearly separable clustering and nonlinearly separable clustering. Due to the high complexity which came from and called as manifold real-world data, nonlinearly separable clustering is one of the problematics in clustering fields has been widely studied and very popular, so because of this part of problematics, it is important to identify nonlinear separable clustering problematics. This study uses both Kernel K-Means and Spectral Clustering Algorithm. This research uses 2 data sets are Two Moon and Two Circles, each dataset tested with both algorithms. The cluster will be evaluated by both internal validation and external validation which each also called as Davies Bouldin Index and Silhouette Index.

*Key Word : Clustering, Kernel K-Means, Spectral Clustering, Davies Bouldin Index, Silhouette Index*

PERBANDINGAN ALGORITMA *CLUSTERING KERNEL K-MEANS* DAN  
*SPECTRAL CLUSTERING* DALAM KASUS *NONLINEAR SEPARABLE*  
*CLUSTERING*

oleh :

Ilham Eldi Aldino  
09021381520061

ABSTRAK

Analisis cluster adalah suatu proses pengelompokan objek, proses ini menghasilkan suatu kelompok yang terdiri dari banyak objek yang juga dapat disebut sebagai cluster, dimana di dalam cluster terdapat objek yang serupa yang memiliki karakteristik sama dengan yang lain dan memiliki karakteristik yang berbeda ke cluster yang berbeda. Mengenai distribusi data cluster, problematika dapat dibagi menjadi dua bagian, itu disebut sebagai linearly separable clustering dan nonlinear separable clustering. Karena kerumitan lipatan yang berasal dari dan disebut sebagai real-world manifold data, pengelompokan nonlinier merupakan salah satu problematika dalam bidang pengelompokan yang telah banyak dipelajari dan sangat populer, sehingga karena bagian problematika ini, penting untuk mengidentifikasi nonlinier yang dapat dipisahkan pengelompokan problematika. Studi ini menggunakan algoritma Kernel K-Means dan algoritma Spectral Clustering. Penelitian ini menggunakan 2 set data yaitu Two Moons dan Two Circles, masing-masing dataset diuji dengan kedua algoritma. Proses hasil clustering akan dievaluasi dengan validasi internal dan validasi eksternal yang masing-masing juga disebut sebagai Davies Bouldin Index dan Silhouette Index.

Kata kunci : *Clustering, Kernel K-Means, Spectral Clusterin, Davies Bouldin Index, Silhouette Index*

## KATA PENGANTAR

### *Bismillahirrahmanir rahim*

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

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan serta dukungan baik secara materi maupun moril. Penulis ingin menyampaikan rasa terimakasih kepada :

1. Orang tuaku, Ir. Elan Ramadanus, S.T dan Mediawaty, S.pd, Saudariku Sabrina Adillah, Diela Andrea, beserta keluarga besarku yang senantiasa mendoakan dan memberikan dukungan baik moril maupun materil.
2. Bapak Jaidan Jauhari, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya, Bapak Rifkie Primartha, M.T selaku ketua jurusan Teknik Informatika dan Ibu Alvi Syahrini Utami, M.Kom selaku Seketaris Jurusan Tekink Informatika.
3. Bapak Danny Matthew Saputra, M.sc selaku Dosen Pembimbing Akademik, yang telah membimbing mengarahkan dan memberikan dukungan penulis dalam proses perkuliahan dan pegerjaan Tugas Akhir.
4. Bapak Rifkie Primartha, M.T selaku Dosen Pembimbing I dan Bapak Danny Matthew Saputra, M.sc selaku Dosen Pembimbing II Tugas Akhir, yang telah mengarahkan dan memotovasi penulis dalam proses perkuliahan dan pengerjaan skripsi.
5. Ibu Dian Palupi Rini, M.Kom., Ph.D selaku Dosen Penguji I dan Bapak Kanda Januar Miraswan, S.Kom., M.T selaku Dosen Penguji II Tugas Akhir yang telah memeberikan masukan dan dorongan dalam penyelesaian skripsi.
6. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Mbak Wiwin dan Kak Rici serta seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
8. Kak Aldy, kak Alkan, kak Apex, kak Irine, kak Lendo, kak Oswald, kak Zuhri, Abiyyu, Adi, Agus, Agung, Aji, Ajrul, Alfredo, Ali, Bondan, Bobby, Gheddi, Hanif, Hasbi, Husein, Imam, Irfan, Mansyah, Rais, Rusdi, Sena, Surya, Ulfi, Vito, Wahyu, Winto, Yusuf serta seluruh teman-teman dan sobat-sobat asemenew jurusan Teknik Informatika lainnya yang telah berbagi keluh kesah, motivasi, dan canda tawa di masa-masa perkuliahan ini.



## DAFTAR ISI

	Halaman
HALAMAN JUDUL .....	i
HALAMAN LEMBAR PENGESAHAN. ....	ii
HALAMAN TANDA LULUS UJIAN SIDANG TUGAS AKHIR .....	iii
HALAMAN PERNYATAAN.....	iv
MOTTO DAN PERSEMBAHAN. ....	v
ABSTRACT. ....	vi
ABSTRAK. ....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI .....	ix
DAFTAR TABEL .....	xiv
DAFTAR GAMBAR.....	xvi
DAFTAR LAMPIRAN.....	xix
<b>BAB I PENDAHULUAN</b>	
1.1 Pendahuluan.....	I-1
1.2 Latar belakang.....	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-11
<b>BAB II LANDASAN TEORI</b>	
2.1 Pendahuluan.....	II-1
2.2 Landasan Teori.....	II-1
2.2.1 Kernel.....	II-1
2.2.1.1 Radial Basis Function Kernel.....	II-3
2.2.2 Graph/Graf.....	II-4

2.2.2.1 Fully Connected Graph.....	I-5
2.2.3 Matrix.....	II-5
2.2.3.1 Laplacian Matrix.....	II-5
2.2.4 Clustering.....	II-6
2.2.4.1 K-Means.....	II-6
2.2.4.2 Kernel K-Means.....	II-9
2.2.4.3 Spectral Clustering.....	II-12
2.2.5 Evaluasi Clustering.....	II-14
2.2.5.1 Davies Bouldin Index.....	II-15
2.2.5.2 Silhouette Index.....	II-16
2.3 Penelitian Yang Relevan.....	II-16
2.3.1 Spectral clustering based on k-nearest neighbors graph.....	II-16
2.3.2 Analisis Perbandingan Metode <i>Fuzzy C-Means</i> dan <i>Subtractive C-Means</i> .....	II-20
2.3.3 <i>Fuzzy C-means based clustering for linearly and nonlinearly separable Data</i> .....	II-21
2.4 Kesimpulan.....	II-26

### BAB III METODOLOGI PENELITIAN

3.1 Pendahuluan.....	III-1
3.2 Data.....	III-1
3.2.1 Jenis, Tipe, dan Sumber Data.....	III-1
3.3 Tahapan Penelitian.....	III-2
3.3.1 Menetapkan Kerangka Kerja / Framework.....	III-2
3.3.2 Menetapkan Kriteria Pengujian.....	III-6
3.3.3 Menentukan Alatyang Digunakan dalam Pelaksanaan Penelitian.....	III-7
3.3.4 Melakukan Pengujian Penelitian.....	III-7
3.3.5 Melakukan Analisa Hasil Pengujian.....	III-8
3.4 Metode Pengembangan Perangkat Lunak.....	III-9
3.4.1 Relational Unified Process.....	III-9
3.4.1.1 Fase Insepsi.....	III-10

3.4.1.2 Fase Elaborasi.....	I-10
3.4.1.3 Fase Konstruksi.....	III-11
3.4.1.4 Fase Transisi.....	III-12
3.5 Manajemen Proyek Penelitian.....	III-12
3.6 Kesimpulan.....	III-17

## BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1Pendahuluan.....	IV-1
4.2 Rational Unified Process.....	IV-1
4.2.1 Fase Insepsi.....	IV-1
4.2.1.1 Pemodelan Bisnis.....	IV-1
4.2.1.2 Kebutuhan Sistem.....	IV-2
4.2.1.3 Analisis dan Desain.....	IV-4
4.2.1.3.1 Analisis Kebutuhan Perangkat Lunak.....	IV-4
4.2.1.3.2 Analisis Data.....	IV-5
4.2.1.3.3 Analisis Algoritma Kernel K-Means.....	IV-5
4.2.1.3.4 Desain Perangkat Lunak.....	IV-6
4.2.2 Fase Elaboasi.....	IV-13
4.2.2.1 Pemodelan Bisnis.....	IV-13
4.2.2.1.1 Perancangan Data.....	IV-13
4.2.2.1.2 Perancangan Interface.....	IV-14
4.2.2.2 Kebutuhan Sistem.....	IV-14
4.2.2.3 Diagram.....	IV-15
4.2.2.3.1 Diagram Activity.....	IV-15
4.2.2.3.2 Diagram Sequence.....	IV-20
4.2.3 Fase Konstruksi.....	IV-21
4.2.3.1 Kebutuhan Sisem.....	IV-22
4.2.3.2 Diagram Class.....	IV-22
4.2.3.3 Kelas Aktivitas.....	IV-22
4.2.3.4 Implementasi.....	IV-23
4.2.3.4.1 Implementasi Kelas.....	IV-24

4.2.3.4.2 Implementasi Antarmuka.....	V-24
4.2.4 Fase Transisi.....	IV-25
4.2.4.1 Pemodelan Bisnis.....	IV-25
4.2.4.2 Kebutuhan Sistem.....	IV-26
4.2.4.3 Rencana Pengujian.....	IV-26
4.2.4.3.1 Rencana Pengujian Use Case Melakukan Proses Clustering Dengan Algoritma Kernel K-Means.....	IV-26
4.2.4.4 Implementasi.....	IV-29
4.2.4.4.1 Pengujian Use Case Dengan Melakukan Proses Menghasilkan Sampel Data Two Moons.....	IV-30
4.2.4.4.2 Pengujian Use Case Dengan Melakukan Proses Menghasilkan Sampel Data Two Circles.....	IV-31
4.2.4.4.3 Pengujian Use Case Dengan Melaksanakan Proses Clustering Dengan Algoritma Kernel K-Means.....	V-33
4.2.4.4.4 Pengujian Use Case Dengan Melaksanakan Proses Clustering Dengan Algoritma Spectral Clustering.....	V-34
4.3 Kesimpulan.....	IV-35

## BAB V HASIL DAN ANALISIS PENELITIAN

5.1 Pendahuluan.....	V-1
5.2 Data Hasil Penelitian.....	V-1
5.2.1 Konfigurasi Pengujian.....	V-1
5.2.2 DataHasil Konfigurasi I.....	V-2
5.2.3 DataHasil Konfigurasi II.....	V-3
5.2.4 DataHasil Konfigurasi III.....	V-4
5.2.5 DataHasil Konfigurasi IV.....	V-5
5.2.6 DataHasil Konfigurasi V.....	V-6
5.3 Analisis Hasil Penelitian.....	V-7
5.4 Kesimpulan.....	V-8

## BAB VI ANALISIS DAN KESIMPULAN

6.1 Pendahuluan.....	I-1
6.2 Kesimpulan.....	VI-1
6.3 Saran.....	VI-2

## DAFTAR PUSTAKA

## DAFTAR TABEL

	Halaman
<b>II-1</b> Hasil perbandingan algoritma <i>clustering</i> <i>Ng et.al.</i> , <i>Fischer et.al.</i> , <i>DaSpec</i> , dan <i>Speclus (Spectral Clustering)</i> .....	I-19
<b>II-2</b> Perbandingan Hasil FCM dan SFCM Data Bangkitan 20.....	II-21
<b>II-3</b> Perbandingan Hasil FCM dan SFCM Data Bangkitan 100.....	II-22
<b>II-4</b> Tabel hasil nilai <i>Information Gain (IG)</i> perbandingan algoritma <i>Neural Gas</i> , <i>Kernel k-means</i> , <i>Kernel Fuzzy c-means</i> , dan <i>Kernel Neural Gas</i> terhadap pengujian setiap data set.....	II-25
<b>II-5</b> Tabel hasil <i>clustering</i> perbandingan algoritma <i>Neural Gas</i> , <i>Kernel k means</i> , <i>Kernel Fuzzy c-means</i> , dan <i>Kernel Neural Gas</i> terhadap pengujian data set <i>wine</i> .....	II-26
<b>III-1</b> Informasi Data Penelitian.....	III-2
<b>III-2</b> Tabel hasil perbandingan <i>clustering</i> algoritma <i>Kernel K-Means</i> dan <i>Spectral Clustering</i> .....	III-8
<b>III-3</b> Tabel Penjadwalan Peneliti dalam Bentuk <i>work Breakdown Structure (WBS)</i> .....	III-16
<b>IV-1</b> Kebutuhan Fungsional.....	IV-3
<b>IV-2</b> Kebutuhan Non-Fungsional.....	IV-4
<b>IV-3</b> Definisi Aktor.....	IV-7
<b>IV-4</b> Definisi <i>Use Case</i> .....	IV-8
<b>IV-5</b> Skenario Proses Menghasilkan <i>Data Set Two Moons</i> .....	IV-9
<b>IV-6</b> Skenario Proses Menghasilkan <i>Data Set Two Circles</i> .....	IV-10
<b>IV-7</b> Skenario Proses <i>Clustering</i> Dengan Algoritma <i>Kernel K-Means</i> .....	IV-11
<b>IV-8</b> Skenario Proses <i>Clustering</i> Dengan Algoritma <i>Spectral Clustering</i> .....	IV-12
<b>IV-9</b> Implementasi Kelas.....	IV-24
<b>IV-10</b> Rencana Pengujian <i>Use Case</i> Melaksanakan Proses Menghasilkan Data	

<i>Set Two Moons</i> .....	V-27
<b>IV-11</b> Rencana Pengujian <i>Use Case</i> Melaksanakan Proses Menghasilkan Data <i>Set Two Circles</i> .....	IV-27
<b>IV-12</b> Rencana Pengujian <i>Use Case</i> Melaksanakan Proses <i>Clustering</i> Dengan Algoritma <i>Kernel K-Means</i> .....	IV-28
<b>IV-13</b> Rencana Pengujian <i>Use Case</i> Melaksanakan Proses <i>Clustering</i> Dengan Algoritma <i>Spectral Clustering</i> .....	IV-28
<b>IV-14</b> Pengujian Proses Menghasilkan Sampel Data <i>Set Two Moons</i> .....	IV-30
<b>IV-15</b> Pengujian Proses Menghasilkan Sampel Data <i>Set Two Circles</i> .....	IV-31
<b>IV-16</b> Pengujian Proses <i>Clustering</i> Dengan Algoritma <i>Kernel K-Means</i> .....	IV-33
<b>IV-17</b> Pengujian Proses <i>Clustering</i> Dengan Algoritma <i>Spectral Clustering</i> .	IV-34
<b>V-1</b> Pengujian Dengan Data <i>Set Two Moons</i> Dengan Jumlah <i>Cluster</i> 1 Sampai Dengan 10 Menggunakan Algoritma <i>Clustering Kernel K-Means</i> .....	V-2
<b>V-2</b> Pengujian Dengan Data <i>Set Two Circles</i> Dengan Jumlah <i>Cluster</i> 1 Sampai Dengan 10 Menggunakan Algoritma <i>Clustering Kernel K-Means</i> .....	V-3
<b>V-3</b> Pengujian Dengan Data <i>Set Two Moons</i> Dengan Jumlah <i>Cluster</i> 1 Sampai Dengan 10 Menggunakan Algoritma <i>Clustering Spectral Clustering</i> .....	V-4
<b>V-4</b> Pengujian Dengan Data <i>Set Two Circles</i> Dengan Jumlah <i>Cluster</i> 1 Sampai Dengan 10 Menggunakan Algoritma <i>Clustering Spectral Clustering</i> .....	V-5
<b>V-6</b> Perbandingan Algoritma <i>Spectral Clustering</i> Dan <i>Kernel K-Means</i> Pada Data <i>Set Two Moons</i> Dan <i>Two Circles</i> .....	V-6

## DAFTAR GAMBAR

	Halaman
<b>II-1</b> Transformasi dimensi data pada 2 kelas <sup>1</sup> .....	I-4
<b>II-2</b> <i>2-cluster run</i> <sup>2</sup> .....	II-7
<b>II-3</b> <i>Two rings datasets two clusters</i> hasil dari <i>kernel k-means</i> <sup>3</sup> .....	II-10
<b>II-4</b> <i>Clustering result</i> .....	II-12
<b>II-5</b> Kinerja/performansi dari pada algoritma <i>spectral clustering</i> (kiri) dan <i>DaSpec</i> (kanan) pterhadap data set buatan ( <i>artificial</i> ) <i>DS1</i> , <i>DS2</i> , <i>DS3</i> , dan <i>DS4</i> .....	II-19
<b>II-6</b> Hasil 2 <i>cluster</i> pada data set <i>Ring</i> .....	II-21
<b>II-7</b> Hasil <i>cluster</i> pada data set <i>Five Cluster</i> .....	II-22
<b>II-8</b> Hasil <i>cluster</i> pada data set <i>Multi-modal Type</i> .....	II-22
<b>II-9</b> Hasil <i>cluster</i> pada data set <i>Gaussian cluster</i> .....	II-23
<b>III-1</b> Penjadwalan Penelitian Tahap Ruang Lingkup dan Unit Penelitian.....	III-17
<b>III-2</b> Penjadwalan Penelitian Tahap Menentukan Dasar Teori yang Berkaitan Dengan Penelitian.....	III-17
<b>III-3</b> Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian.....	III-17
<b>III-4</b> Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian.....	III-18
<b>III-5</b> Penjadwalan Penelitian Tahap Melakukan Pengujian Penelitian dan Membuat Analisa Hasil Pengujian dan Membuat Kesimpulan.....	III-18
<b>IV-1</b> Diagram Use Case.....	IV-7
<b>IV-2</b> Desain Interface.....	IV-14
<b>IV-3</b> Diagram Aktivitas Proses Menghasilkan/Meng-generate Data Set Two Moons.....	IV-16
<b>IV-4</b> Diagram Aktivitas Proses Menghasilkan/Meng-generate Data Set Two	



Circles.....	V-17
<b>IV-5</b> Diagram Aktivitas Proses Clustering Dengan Algoritma Kernel K-Means.....	IV-18
<b>IV-6</b> Diagram Aktivitas Proses Clustering Dengan Algoritma Spectral Clustering.....	IV-19
<b>IV-7</b> Diagram Sequence Algoritma Kernel K-Means.....	IV-20
<b>IV-8</b> Diagram Sequence Algoritma Spectral Clustering.....	IV-21
<b>IV-9</b> Diagram Kelas.....	IV-22
<b>IV-10</b> Kelas Analisis Menjalankan Proses Clustering Dengan Algoritma Kernel K-Means.....	IV-23
<b>IV-11</b> Kelas Analisis Menjalankan Proses Clustering Dengan Algoritma Spectral Clustering.....	IV-23
<b>IV-12</b> Tampilan Interface Halaman Utama Piranti Lunak.....	IV-25
<b>V-1</b> Data Set Two Moons Hasil Proses Clustering Kenrel K-Means Dengan Jumlah Cluster 1 Dan 2.....	V-13
<b>V-2</b> Data Set Two Moons Hasil Proses Clustering Kenrel K-Means Dengan Jumlah Cluster 3 Dan 4.....	V-13
<b>V-3</b> Data Set Two Moons Hasil Proses Clustering Kenrel K-Means Dengan Jumlah Cluster 5 Dan 6.....	V-13
<b>V-4</b> Data Set Two Moons Hasil Proses Clustering Kenrel K-Means Dengan Jumlah Cluster 7 Dan 8.....	V-14
<b>V-5</b> Data Set Two Moons Hasil Proses Clustering Kenrel K-Means Dengan Jumlah Cluster 9 Dan 10.....	V-14
<b>V-6</b> Data Set Two Circles Hasil Proses Clustering Kenrel K-Means Dengan Jumlah Cluster 1 Dan 2.....	V-14
<b>V-7</b> Data Set Two Circles Hasil Proses Clustering Kenrel K-Means Dengan Jumlah Cluster 3 Dan 4.....	V-15
<b>V-8</b> Data Set Two Circles Hasil Proses Clustering Kenrel K-Means Dengan Jumlah Cluster 5 Dan 6.....	V-15
<b>V-9</b> Data Set Two Circles Hasil Proses Clustering Kenrel K-Means Dengan Jumlah Cluster 7 Dan 8.....	V-15

<b>V-10</b> Data Set Two Circles Hasil Proses Clustering Kenrel K-Means Dengan Jumlah Cluster 9 Dan 10 .....	V-16
<b>V-11</b> Data Set Two Moons Hasil Proses Clustering Spectral Clustering Dengan Jumlah Cluster 1 Dan 2.....	V-16
<b>V-12</b> Data Set Two Moons Hasil Proses Clustering Spectral Clustering Dengan Jumlah Cluster 3 Dan 4.....	V-16
<b>V-13</b> Data Set Two Moons Hasil Proses Clustering Spectral Clustering Dengan Jumlah Cluster 5 Dan 6.....	V-17
<b>V-14</b> Data Set Two Moons Hasil Proses Clustering Spectral Clustering Dengan Jumlah Cluster 7 Dan 8.....	V-17
<b>V-12</b> Data Set Two Moons Hasil Proses Clustering Spectral Clustering Dengan Jumlah Cluster 9 Dan 10.....	V-17
<b>V-13</b> Data Set Two Circles Hasil Proses Clustering Spectral Clustering Dengan Jumlah Cluster 1 Dan 2.....	V-18
<b>V-14</b> Data Set Two Circles Hasil Proses Clustering Spectral Clustering Dengan Jumlah Cluster 3 Dan 4.....	V-18
<b>V-15</b> Data Set Two Circles Hasil Proses Clustering Spectral Clustering Dengan Jumlah Cluster 5 Dan 6.....	V-18
<b>V-16</b> Data Set Two Circles Hasil Proses Clustering Spectral Clustering Dengan Jumlah Cluster 7 Dan 8.....	V-19
<b>V-15</b> Data Set Two Circles Hasil Proses Clustering Spectral Clustering Dengan Jumlah Cluster 9 Dan 10.....	V-19

## DAFTAR LAMPIRAN

1. <i>Source Code</i> Program.....	L-2
------------------------------------	-----

# BAB I

## PENDAHULUAN

### 1.1 Pendahuluan

Pada bab ini akan menyinggung tentang pembahasan mengenai gambaran penelitian secara garis besar, yaitu latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan, dan kesimpulan.

### 1.2 Latar Belakang

Analisis *cluster* merupakan proses pengelompokkan objek-objek, proses tersebut menghasilkan sebuah kelompok yang terdiri dari banyak objek yang dapat disebut dengan *cluster*, yang di dalamnya terdapat objek-objek yang serupa/memiliki kesamaan satu dengan yang lainnya dan berbeda terhadap *cluster* yang lainnya (Soni & Ganatra, 2012). Dalam arti lain, *clustering* diketahui sebagai bentuk klasifikasi tanpa pengawasan (*unsupervised classification*), yaitu tindakan mempartisi pola-pola data dalam bentuk sub-kelompok sehingga pola tersebut ter-*cluster* secara bersamaan (Saxena, 2017).

Mengenai perihal tentang distribusi data *cluster*, permasalahan pada *data clustering* dapat dikategorikan menjadi 2 yaitu *linearly separable clustering* dan

*nonlinearly separable clustering*. Dikarenakan sebuah lipatan kompleks yang berasal dari dan disebut sebagai *manifold real-world data*, *nonlinearly separable clustering* merupakan satu diantara dari permasalahan di dalam *clustering* yang paling populer dan secara luas telah dikaji. *Nonlinear separable clustering* dapat didefinisikan sebagai bagian permasalahan dalam *cluster distribution* yang paling sedikit mengandung 2 *cluster* dengan melibatkan data yang berbentuk cekung dan tidak rapi atau bisa disebut sebagai *arbitrary shape*. Dikarenakan timbulnya istilah *manifold real-world data*, maka penting untuk mengidentifikasi *nonlinear separable cluster*. Bagaimanapun juga metode klasik *linear separable clustering* seperti *k-means* tidak dapat memenuhi kebutuhan dalam identifikasi *nonlinear cluster* (Wang & Lai, 2016). Oleh sebab itu, *nonlinear separable clustering* memerlukan teknik *clustering* yang berbeda dari *linear separable clustering*.

Penelitian terdahulu tentang algoritma *nonlinearly separable clustering* telah dikaji sebelumnya, Penelitian tersebut membandingkan algoritma *spectral clustering* dengan 3 algoritma lain yang berbeda yaitu Ng *et al*, Fischer *et al*, dan *DaSpec*, yang masing-masing diuji menggunakan 6 *artificial data set* dan 3 *real-world problem data set* (Lucinska & Wierzchon, 2012). Berdasarkan dari hasil penelitian yang telah dilakukan, algoritma *spectral clustering* memberikan hasil yang paling baik diantaranya dengan memiliki kemampuan paling fleksibel dan mampu berkinerja baik secara independen terhadap kumpulan struktur data dengan memberikan *number of incorrectly clustered points* terbilang jauh lebih kecil diantara yang lain.

Penelitian yang berhubungan lainnya adalah pengembangan algoritma *Neutrosophic c-Means* menjadi *Kernel Neutrosophic c-Means* (KNCM) menggunakan 3 jenis data set diantaranya *artificial datasets*, *real datasets*, dan *image segmentation*. Kesimpulannya pada pengujian algoritma KNCM cukup berhasil pada proses *clustering* beberapa jenis *toy dataset* dan *real datasets*, selain karena efisiensinya, algoritma KNCM mampu mengatasi data yang berderau dan data pencilan (Akbulut et al., 2017).

Penelitian terdahulu yang serupa juga dilakukan dengan mengembangkan algoritma *Hartigan's* menjadi kombinasi antara *kernel spaces* dan algoritma yang diusulkan yang disebut sebagai *kernel k-groups*. Algoritma *kernel k-groups* telah dibandingkan dengan *kernel k-means* dan *spectral clustering* oleh peneliti. Menurut angka dari hasil perhitungan, algoritma *kernel k-group* menampilkan performansi yang unggul diantaranya. Peneliti menekankan bahwa algoritma *kernel k-group* memiliki kompleksitas yang sama dengan algoritma *kernel k-means*. Akan tetapi algoritma *kernel k-group* memiliki keterbatasan yang sama dengan algoritma *spectral clustering* dan *kernel k-means* yang mengelompokkan data secara tidak seimbang (Franca, Rizzo, & Vogelstein, 2018 ).

Berlandaskan penelitian terdahulu yang pernah dilakukan sebelumnya, masing-masing algoritma yang berperan dalam mengatasi salah satu permasalahan dalam *clustering* yaitu *nonlinear separable clustering* memiliki kualitas performansi, teknik, kompleksitas komputasi, dan studi kasus yang berbeda-beda, untuk itu penulis akan melakukan penelitian tentang

perbandingan algoritma *spectral clustering* dan *kernel k-means*.

### **11 Rumusan Masalah**

Rumusan masalah pada penelitian ini adalah algoritma manakah yang memiliki kemampuan paling baik diantara algoritma *Spectral Clustering* (SC) dan *Kernel K-Means* (KKM) untuk penyelesaian masalah dalam kasus *nonlinear separable clustering*.

Untuk memecahkan rumusan masalah tersebut maka dibentuk menjadi 4 pertanyaan penelitian :

1. Bagaimana kemampuan *clustering* menggunakan algoritma *spectral clustering* (SC) dan *Kernel K-Means* (KKM) dalam kasus pengelompokan *Nonlinear Separable Clustering* ?
2. Algoritma manakah diantara *Kernel K-Means* dan *Spectral Clustering* yang memiliki kinerja terbaik dalam menangani kasus *Nonlinear Separable Clustering*?

### **12 Tujuan Penelitian**

Beberapa tujuan dari penelitian ini adalah :

1. Mengetahui kemampuan masing-masing algoritma *Kernel K-Means* dan *Spectral Clustering* dalam menangani kasus *nonlinear separable cluster* pada data set *Two Moons* dan *Two Circles*.

2. Mengetahui algoritma yang memiliki kemampuan terbaik dalam mengatasi kasus nonlinear separable cluster pada data set *Two Moons* dan *Two Circles*.

### 13 Manfaat Penelitian

Manfaat yang diharapkan dari pengerjaan karya ilmiah dan pada penelitian ini diantaranya adalah :

1. Mempelajari tentang teknik *clustering nonlinear separable clustering*.
2. Membantu penelitian setelahnya sebagai referensi yang berkaitan terhadap penelitian yang berhubungan dengan teknik *clustering nonlinear separable clustering*.

### 14 Batasan Masalah

Batasan masalah pada riset ini diantaranya adalah sebagai berikut.

Batasan Masalah :

1. Algoritma *clustering* yang digunakan sebagai perbandingan pada penelitian ini adalah algoritma *Kernel K-Means* dan *Spectral Clustering (Unnormalized Spectral Clustering)*
2. Metode Kernel yang digunakan pada masing-masing algoritma adalah metode kernel RBF (*Radial Basis Function*).
3. Besaran parameter *gamma* ( $\sigma$ ) pada metode kernel RBF yang digunakan pada penelitian ini ditentukan secara *trial and error*.



4. Jumlah cluster/nilai K pada penelitian ini ditentukan secara *pre-defined* (ditentukan sebelum di lakukannya proses clustering) yang berjumlah 1 hingga 10 cluster untuk masing-masing data set.
5. Pengukuran perbandingan kemampuan masing-masing algoritma pada penelitian ini menggunakan *Davies Bouldin Index* atau disebut juga sebagai DBI (Validasi Internal) dan *Silhouette* (Validasi Eksternal).
6. Data yang digunakan sebagai objek penelitian ini merupakan *data set Two Moons* dan *Two Circles*

## **15 Sistematika Penulisan**

### **BAB I. PENDAHULUAN**

Pembahasan yang akan disampaikan pada bab I adalah latar belakang masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan dan konklusi/kesimpulan.

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

Penjelasan yang akan disampaikan pada bab II adalah pokok-pokok teori yang dimanfaatkan pada penelitian ini, penjabaran pada bab dua menyinggung tentang perolehan/hasil dari penelitian terdahulu yang pernah dilakukan sebelumnya, cara pengelompokkan *datasets* dalam kasus *nonlinear separable clustering*, penjabaran tentang algoritma *Spectral Clustering* (SC) dan *Kernel K-Means* (KKM), dan pemaparan lainnya yang berhubungan dengan penelitian.

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

Pembahasan yang akan disampaikan pada bab III adalah tahap-tahap yang akan dieksekusi dalam penelitian ini. Setiap rancangan tahap-tahap penelitian digambarkan secara terperinci yang mengarah pada kerangka kerja. Pada penghabisan bab III ini memuat perancangan manajemen proyek pada implementasi penelitian.

#### **1.6 Kesimpulan**

Pengelompokkan dalam kasus *nonlinear separable clustering* berperan penting dalam pengelompokkan *dataset* secara *nonlinear grouping* dalam disiplin ilmu *clustering*. Dikarenakan sebab itu, penelitian ini akan menguraikan hasil perbandingan kemampuan kinerja masing-masing algoritma *spectral clustering* (SC) dan *kernel k-means* (KKM). Objek yang dilibatkan pada penelitian ini adalah *artificial data set/data set dummy* yang dihasilkan oleh *library python*.

## DAFTAR PUSTAKA

Akbulut, Y., Senegur, A., Guo, Y., & Polat, K. (2017). KNCM: Kernel Neutrosophic c-Means Clustering. *Applied Soft Computing*, 52, 714-724.

Camastra, F., & Verri, A. (n.d.). A Novel Kernel Method for Clustering. *Biological and Artificial Intelligence Enviroments*, 245-250.

Franca, G., Vogelstein, J. T., & Rizzo, M. (2020). Kernel K-Groups via Hartigan's Method. *IEEE Transaction of Pattern Analysis and Machine Intelligence*, 1-1.

Jain, A. K. (2010). Data Custering: 50 years beyond K-Means. *Pattern Recognition Letters*, 31(8), 651-666.

Lucinska, M., & Wierzchon, S. T. (2012). Spectral Clustering Based on k-Nearest Neighbor Graph. *International Federation for Information Processing*, 254-265.

Pahwa, S., Sinwar, D. (2015). Comparison Of Various Kernels Of Support Vector Machine. *International Journal for Research in Applied Science & Engineering Technology*, 3(7), 2321-9653..

Wang, C-. D., Lai, J. -H. (2016). *Nonlinear Clustering: Methods and*