

**Penentuan Pusat Kluster Secara Otomatis Pada Algoritma  
Density Peaks Clustering Berbasis Metode Inter Quartile Range**

**SKRIPSI**



**Oleh:**

**Ridwan Efendi**

**20081010071**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"**

**JAWA TIMUR**

**2024**

# LEMBAR PENGESAHAN

## SKRIPSI

Judul : Penentuan Pusat Kluster Secara Otomatis Pada Algoritma Density  
Peaks Clustering Berbasis Metode Inter Quartile Range

Oleh : Ridwan Efendi

NPM : 20081010071

Telah Diseminarkan Dalam Ujian Skripsi Pada :

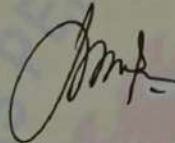
Hari Kamis, Tanggal 04 Juli 2024

### Mengetahui

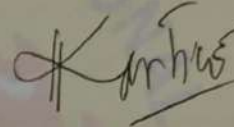
Dosen Pembimbing

Dosen Penguji

1.



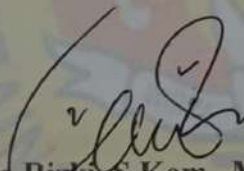
1.



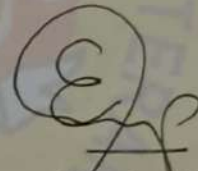
Achmad Junaidi, S.Kom, M.Kom  
NPT : 3 7811 04 0199 1

Dr. Ir. Kartini, S.Kom. MT.  
NIP : 19611110 199103 2 001

2.



2.



Agung Mustika Rizki, S.Kom., M.Kom  
NIP : 19930725 202203 1 008

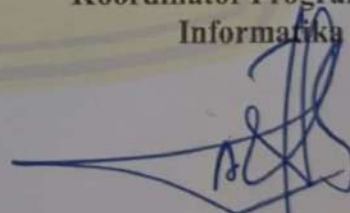
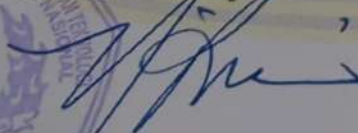
Eka Prakarsa Mandyartha, S.T., M.Kom  
NIP : 19880525 201803 1 001

### Menyetujui

Dekan

Koordinator Program Studi  
Informatika

Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT.  
NIP : 19681126 199403 2 001

Fetty Tri Anggraeny, S.Kom, M.Kom  
NIP : 19820211 202121 2 005

## SURAT PERNYATAAN BEBAS DARI PLAGIASI

Saya, mahasiswa Program Studi Sarjana Informatika Universitas Pembangunan Nasional "Veteran" Jawa Timur, yang bertanda tangan di bawah ini:

Nama : Ridwan Efendi

NPM : 20081010071

Menyatakan dengan sesungguhnya bahwa Skripsi/Tugas Akhir yang saya kerjakan berjudul:

"Penentuan Pusat Klaster Secara Otomatis Pada Algoritma Density Peaks Clustering Berbasis Metode Inter Quartile Range"

bukan merupakan plagiasi sebagian atau keseluruhan dari Skripsi/Tugas Akhir/Penelitian orang lain dan juga bukan merupakan produk dan software yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/Tugas Akhir ini secara keseluruhan adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional "Veteran" Jawa Timur maupun di Institut Pendidikan lain. Bukti hasil pengecekan plagiasi dokumen ini dapat ditelusuri melalui QR Code di bawah.

Apabila di kemudian hari terbukti bahwa dokumen ini merupakan plagiasi karya orang lain, saya sanggup menerima sanksi sesuai aturan yang berlaku.

Demikian atas perhatiannya disampaikan terima kasih.

Surabaya, 04 Juli 2024

Hormat saya,



Ridwan Efendi

NPM. 20081010071

# **Penentuan Pusat Kluster Secara Otomatis Pada Algoritma Density Peaks Clustering Berbasis Metode Inter Quartile Range**

**Nama Mahasiswa : Ridwan Efendi**

**NPM : 20081010071**

**Program Studi : Teknik Informatika**

**Dosen Pembimbing : Achmad Junaidi, S.Kom, M.Kom**

**Agung Mustika Rizki, S.Kom., M.Kom**

## **Abstrak**

*Clustering* adalah sebuah metode untuk mengelompokkan data yang sejenis ke dalam satu bagian yang sama. Proses ini mampu membantu manusia untuk mendapatkan informasi secara lebih cepat. Dalam konteks media sosial misalnya, metode *clustering* dapat memberikan informasi terkait konten yang cenderung disukai dan kurang disukai. Algoritma *Density Peaks Clustering (DPC)* adalah salah satu algoritma yang cukup populer digunakan untuk mengelompokkan sebuah data. Sudah banyak penelitian yang menggunakan algoritma ini. Namun, algoritma DPC memiliki kekurangan dalam hal penentuan pusat kluster. Pusat Kluster dalam algoritma DPC masih dipilih secara manual melalui grafik keputusan. Pemilihan pusat kluster secara otomatis menambah subjektivitas dan ketidakstabilan dalam algoritma. Untuk mengatasi masalah tersebut, digunakan sebuah algoritma atau metode ‘Penentuan Pusat Otomatis’ yang berbasis pada metode *Inter Quartile Range (IQR)*. Dalam penelitian ini, gabungan metode DPC dan ‘Penentuan Pusat Otomatis’ disebut dengan algoritma DPC-IQRSM. Dalam penelitian ini, dilakukan uji coba untuk mengevaluasi keefektifan algoritma DPC-IQRSM menggunakan dataset berdimensi rendah seperti *iris*, *aggregation*, *flame*, dan *spiral*. Selain itu, dilakukan perbandingan dengan dua algoritma *clustering* lainnya, yaitu K-means

dan GB-DPC (*Gab Based Density Peak Clustering*). Hasil evaluasi menunjukkan algoritma DPC-IQRSM lebih baik dibandingkan dengan algoritma lainnya.

***Kata Kunci: Clustering, Density Peaks Clustering, Deteksi Pusat Otomatis***

# **Penentuan Pusat Kluster Secara Otomatis Pada Algoritma Density Peaks Clustering Berbasis Metode Inter Quartile Range**

**Nama Mahasiswa : Ridwan Efendi**

**NPM : 20081010071**

**Program Studi : Teknik Informatika**

**Dosen Pembimbing : Achmad Junaidi, S.Kom, M.Kom**

**Agung Mustika Rizki, S.Kom., M.Kom.**

## **Abstract**

Clustering is a method for grouping similar data into the same part. This process helps humans to obtain information more quickly. In the context of social media, for example, clustering methods can provide insights into content preferences and dislikes. The Density Peaks Clustering (DPC) algorithm is a popular choice for data clustering. Many studies have utilized this algorithm. However, DPC algorithm suffers from shortcomings in determining cluster centers. Cluster centers in DPC are still manually selected through decision graphs, introducing subjectivity and instability. To address this issue, an algorithm or method called 'Automatic Center Determination' based on the Inter Quartile Range (IQR) method is employed. In this study, the combination of DPC and Automatic Center Determination is referred to as DPC-IQRSM algorithm. The research evaluates the effectiveness of the DPC-IQRSM algorithm using low-dimensional datasets such as iris, aggregation, flame, and spiral. Additionally, comparisons are made with two other clustering algorithms, K-means and GB-DPC (Gab Based Density Peak Clustering). The evaluation results indicate that the DPC-IQRSM algorithm outperforms the others.

***Kata Kunci: Clustering, Density Peaks Clustering, Deteksi Pusat Otomatis***

## **KATA PENGANTAR**

Puji syukur kita panjatkan kehadiran Allah Swt. yang telah memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan laporan skripsi yang berjudul “Penentuan Pusat Klaster Secara Otomatis Pada Algoritma Density Peaks Clustering Berbasis Metode Inter Quartile Range” ini tepat pada waktunya. Penulis berusaha dengan semaksimal mungkin dalam penyusunan laporan ini baik dari hasil bimbingan bersama dosen pembimbing, maupun diskusi dengan teman-teman penulis.

Surabaya, 04 Juli 2024

Penulis,

Ridwan Efendi

NPM. 20081010071

## UCAPAN TERIMA KASIH

Dalam kesempatan ini, penulis mengucapkan banyak terima kasih kepada semua pihak yang telah membantu dalam pelaksanaan kegiatan perkuliahan maupun penyusunan laporan ini kepada:

1. Tuhan Yang Maha Esa, yang senantiasa melimpahkan berkat dan rahmat kepada penulis.
2. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT., selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT., selaku Dekan Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom. selaku Koordinator Program Studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Bapak Andreas Nugroho Sihananto, S.Kom., M.Kom., selaku Koordinator Skripsi Program Studi Informatika Universitas Pembangunan “Veteran” Jawa Timur.
6. Bapak Achmad Junaidi, S.Kom, M.Kom., selaku dosen pembimbing 1 yang telah membimbing penulis dari awal penyusunan laporan sampai penandatanganan laporan ini.
7. Bapak Agung Mustika Rizki, S.Kom., M.Kom., selaku dosen pembimbing 2 yang telah membimbing penulis dari awal penyusunan laporan sampai penandatanganan laporan ini.
8. Ibu Dr. Ir. Kartini, S.Kom. MT., selaku dosen penguji 1 yang telah memberikan masukan dan kritik yang membangun selama sidang skripsi maupun masa revisi skripsi.
9. Bapak Eka Prakarsa Mandyartha, S.T., M.Kom., selaku dosen penguji 2 yang telah memberikan masukan dan kritik yang membangun selama sidang maupun masa revisi skripsi.
10. Ibu Henni Endah Wahanani, ST. M.Kom. selaku dosen wali Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.



11. Seluruh dosen dan staff Program Studi Informatika dan staff Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
12. Teristimewa Kedua Orang Tua yang sangat penulis cintai, Ayahanda Meseran dan Ibunda Sumarni, yang telah membantu memberikan semangat, doa, kasih harapan dan dorongan moral dan spiritual kepada penulis dalam menyelesaikan tugas akhir skripsi.
13. Kakak tercinta Indah Puspitasari dan Kusnul Patimah, terima kasih banyak atas dukungan, kasih dan doanya selalu untuk penulis untuk mencapai impian.
14. Untuk sahabat kuliah Dimas Octa Maulana, Rico Putra Anugerah, dan Achmad Fauzihan Bagus Sajiwo terima kasih atas segala pengalaman, kebersamaan, semangat, saran, dan doa yang diberikan sehingga penulis dapat menyelesaikan tugas akhir skripsi.
15. Teman-teman penulis lainnya yang senantiasa bersama dari semester 1 sampai sekarang.

Akhir kata, penulis berharap laporan ini dapat memberikan manfaat serta menjadi bahan referensi bagi penulis maupun pembaca. Penulis menyadari bahwa laporan yang ditulis ini masih jauh dari kata sempurna. Oleh karena itu, kritik dan saran yang membangun penulis butuhkan demi masa depan bersama.

## DAFTAR ISI

LEMBAR PENGESAHAN.....	ii
SURAT PERNYATAAN BEBAS DARI PLAGIASI .....	iii
Abstrak.....	iv
Abstract .....	vi
KATA PENGANTAR .....	vii
UCAPAN TERIMA KASIH.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL .....	xvi
BAB I.....	1
PENDAHULUAN .....	1
1.1    Latar belakang .....	1
1.2    Rumusan Masalah.....	4
1.3    Tujuan .....	4
1.4    Manfaat .....	5
1.5    Batasan Masalah.....	5
BAB II.....	6
TINJAUAN PUSTAKA.....	6
2.1    Penelitian Terdahulu .....	6
2.2 <i>Density Peaks Clustering (DPC)</i> .....	10
2.3    Kuartil ( <i>Quartile</i> ).....	13
2.4    Metode <i>Inter Quartile Range (IQR)</i> .....	15

2.5	<i>Adjusted Rand Index (ARI)</i> .....	15
2.6	<i>Normalized Mutual Information (NMI)</i> .....	16
2.7	Python .....	16
2.8	Google Collaboratory .....	18
BAB III.....		20
METODE PENELITIAN.....		20
3.1	Alur Penelitian.....	20
3.2	Studi Literatur.....	20
3.3	Pengumpulan Data.....	21
3.4	Preprocessing Data .....	22
3.4.1	<i>Data Cleaning</i> .....	23
3.4.2	Label Encoding.....	26
3.5	Proses <i>Clustering</i> .....	28
3.6	Penentuan Pusat Kluster Secara Otomatis .....	36
3.7	Evaluasi.....	38
BAB IV .....		39
HASIL DAN PEMBAHASAN .....		39
4.1	Pengumpulan Data.....	39
4.1.1	Dataset Iris .....	39
4.1.2	Dataset Aggregation .....	41
4.1.3	Dataset Flame .....	42
4.1.4	Dataset Spiral .....	43
4.2	Implementasi Program.....	44

4.2.1	<i>Library</i> .....	44
4.2.2	Implementasi <i>Preprocessing</i> Data.....	44
4.2.3	Implementasi Menghitung Jarak Euclidean .....	46
4.2.4	Implementasi Menghitung <i>Cutoff distance</i> .....	48
4.2.5	Implementasi Menghitung Kepadatan Lokal .....	51
4.2.6	Implementasi Menghitung Jarak Lokal .....	55
4.2.7	Implementasi Membuat Grafik Keputusan .....	59
4.2.8	Implementasi Proses Pengelompokan .....	62
4.2.9	Algoritma Penentuan Pusat Kluster Otomatis.....	63
4.3	Pembahasan Pengujian.....	67
BAB V.....		83
KESIMPULAN DAN SARAN .....		83
5.1	Kesimpulan .....	83
5.2	Saran .....	83
Daftar Pustaka .....		85

## DAFTAR GAMBAR

Gambar 1.1 Titik Pusat .....	3
Gambar 2.1 DPC dalam Grafik dua Dimensi. (A) Distribusi data. (B) Grafik Keputusan.....	12
Gambar 3.1 Diagram Alir Penelitian .....	20
Gambar 3.2 Diagram Alir Preprocessing Data.....	22
Gambar 3.3 Grafik keputusan untuk data dari Tabel 3.10 .....	34
Gambar 3.4 Gambar grafik keputusan 1 dimensi dan 2 dimensi.....	36
Gambar 4.1 Dataset iris.....	40
Gambar 4.2 Bar plot jumlah setiap kategori pada dataset iris.....	40
Gambar 4.3 Bar plot jumlah setiap label pada dataset aggregation.....	41
Gambar 4.4 Bar plot jumlah setiap label pada dataset flame .....	42
Gambar 4.5 Bar plot jumlah setiap label pada dataset spiral .....	43
Gambar 4.6 Matriks jarak euclidean dataset iris .....	46
Gambar 4.7 Matriks jarak euclidean dataset aggregation .....	47
Gambar 4.8 Matriks jarak euclidean dataset flame .....	47
Gambar 4.9 Matriks jarak euclidean dataset spiral.....	48
Gambar 4.10 Nilai cutoff distance untuk dataset iris.....	48
Gambar 4.11 Nilai cutoff distance untuk dataset aggregation .....	49
Gambar 4.12 Nilai cutoff distance untuk dataset flame.....	49
Gambar 4.13 Nilai cutoff distance untuk dataset spiral.....	49
Gambar 4.14 Grafik keputusan dataset iris .....	60
Gambar 4.15 Grafik keputusan dataset aggregation.....	61
Gambar 4.16 Grafik keputusan dataset flame .....	61

Gambar 4.17 Grafik keputusan dataset spiral .....	62
Gambar 4.18 Grafik Keputusan.....	64
Gambar 4.19 Gambar untuk ilustrasi menghitung konektivitas .....	65
Gambar 4.20 Pemotongan konektivitas .....	67
Gambar 4.21 Kebenaran klaster dari dataset iris .....	68
Gambar 4.22 Hasil <i>clustering</i> dataset iris menggunakan k-means.....	68
Gambar 4.23 Hasil clustering dataset iris menggunakan DPC .....	69
Gambar 4.24 Hasil <i>clustering</i> dataset iris menggunakan GB-DPC.....	70
Gambar 4.25 Hasil <i>clustering</i> dataset iris menggunakan DPC-IQRSM.....	71
Gambar 4.26 Kebenaran klaster dari dataset aggregation.....	72
Gambar 4.27 Hasil <i>clustering</i> algoritma k-means pada dataset aggregation .....	72
Gambar 4.28 Hasil <i>clustering</i> algoritma DPC pada dataset aggregation.....	73
Gambar 4.29 Hasil <i>clustering</i> algoritma GB-DPC untuk dataset aggregation .....	74
Gambar 4.30 Hasil <i>clustering</i> algoritma DPC-IQRSM untuk dataset aggregation .....	75
Gambar 4.31 Kebenaran klaster dari dataset flame .....	75
Gambar 4.32 Hasil <i>clustering</i> algoritma k-means pada dataset flame .....	76
Gambar 4.33 Hasil <i>clustering</i> algoritma DPC pada dataset flame .....	76
Gambar 4.34 Hasil <i>clustering</i> algoritma GB-DPC pada dataset flame .....	77
Gambar 4.35 Hasil <i>clustering</i> algoritma DPC-IQRSM dataset flame.....	77
Gambar 4.36 Kebenaran klaster dataset spiral .....	78
Gambar 4.37 Hasil <i>clustering</i> algoritma K-means untuk dataset spiral .....	79
Gambar 4.38 Hasil <i>clustering</i> algoritma DPC untuk dataset spiral.....	79
Gambar 4.39 Hasil <i>clustering</i> algoritma GB-DPC untuk dataset spiral .....	80

Gambar 4.40 Hasil *clustering* algoritma DPC-IQRSM untuk dataset spiral ..... 80

## DAFTAR TABEL

Tabel 3.1 Dataset Penelitian.....	21
Tabel 3.2 Contoh data yang tersimpan pada dataset.....	21
Tabel 3.3 Contoh dataset dengan missing value.....	24
Tabel 3.4 Contoh dataset tanpa missing value .....	24
Tabel 3.5 Contoh duplikasi data.....	25
Tabel 3.6 Contoh dataset tanpa missing value dan duplikasi data .....	26
Tabel 3.7 Contoh label dari sebuah dataset.....	27
Tabel 3.8 Contoh fitur dari sebuah dataset.....	27
Tabel 3.9 Contoh hasil proses label encoding .....	28
Tabel 3.10 Data untuk contoh perhitungan menggunakan rumus-rumus pada algoritma DPC.....	29
Tabel 3.11 Jarak euclidean distance untuk setiap titik data pada Tabel 3.10.....	30
Tabel 3.12 Nilai kepadatan lokal untuk setiap titik data pada Tabel 3.10 .....	32
Tabel 3.13 Nilai jarak lokal untuk setiap titik data pada Tabel 3.10 .....	33
Tabel 3.14 Hasil sorting data pada Tabel 3.12 .....	35
Tabel 3.15 Contoh hasil pengelompokan.....	36
Tabel 3.16 Parameter untuk pengujian .....	38
Tabel 4.1 Dataset aggregation .....	41
Tabel 4.2 Contoh dataset flame .....	42
Tabel 4.3 Contoh dataset spiral .....	43
Tabel 4.4 Perubahan jumlah data.....	45
Tabel 4.5 Contoh nilai kepadatan pada dataset iris .....	52
Tabel 4.6 Contoh nilai kepadatan pada dataset aggregation .....	53



Tabel 4.7 Contoh nilai kepadatan pada dataset flame.....	54
Tabel 4.8 Contoh nilai kepadatan pada dataset spiral.....	55
Tabel 4.9 Contoh nilai jarak lokal pada dataset iris.....	56
Tabel 4.10 Contoh nilai jarak lokal pada dataset aggregation .....	57
Tabel 4.11 Contoh nilai jarak lokal pada dataset flame .....	58
Tabel 4.12 Contoh nilai jarak lokal pada dataset spiral .....	59
Tabel 4.13 Rangkuman hasil kualitas clustering .....	81