

TUGAS AKHIR

**PENENTUAN *CLUSTER* OPTIMUM UNTUK
MENGELOMPOKKAN KABUPATEN DI JAWA TIMUR
BERDASARKAN TINGKAT PENGANGGURAN DAN TINGKAT
KEMISKINAN MENGGUNAKAN METODE *K-MEANS***

Disusun Untuk Melengkapi Dan Memenuhi Syarat Kelulusan
Guna Meraih Gelar Sarjana Komputer
Teknik Informatika Universitas Muhammadiyah Jember



Eva Alfariani
1510651073

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER
2019**

HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR

PENENTUAN *CLUSTER* OPTIMUM UNTUK

MENGELOMPOKAN KABUPATEN DI JAWA TIMUR

BERDASARKAN TINGKAT PENGANGGURAN DAN TINGKAT

KEMISKINAN MENGGUNAKAN METODE *K-MEANS*

Oleh :
Eva Alfariani
1510651073

Telah disetujui bahwa Laporan Tugas Akhir ini untuk diajukan pada sidang Tugas Akhir sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom) di Universitas Muhammadiyah Jember

Disetujui oleh,

Pembimbing I



Ari Eko Wardoyo, S.T, M.Kom
NIP. 1975 0214 2005 0110 01

Pembimbing II



Agung Nilogiri, S.T, M.Kom
NIP. 19770330 200501 1 002

HALAMAN PENGESAHAN

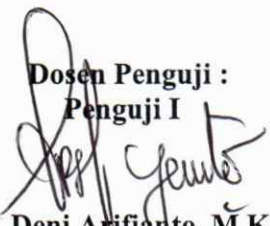
PENENTUAN *CLUSTER* OPTIMUM UNTUK MENGELOMPOKAN KABUPATEN DI JAWA TIMUR BERDASARKAN TINGKAT PENGANGGURAN DAN TINGKAT KEMISKINAN MENGGUNAKAN METODE *K-MEANS*

Oleh :
Eva Alfariani
1510651073

Telah mempertanggung jawabkan Laporan Tugas Akhirnya pada sidang Tugas Akhir tanggal 19-07-2019 sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom) di Universitas Muhammadiyah Jember

Disetujui oleh,

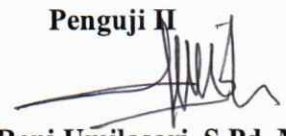
Dosen Penguji :
Penguji I


Deni Arifianto, M.Kom
NIDN. 0718068103

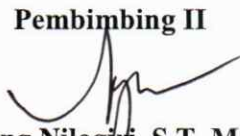
Dosen Pembimbing :
Pembimbing I


Ari Eko Wardoyo, S.T, M.Kom
NIP. 1975 0214 2005 0110 01

Penguji II


Reni Umilasari, S.Pd, M.Si
NIDN. 0728079101

Pembimbing II


Agung Nilogiri, S.T, M.Kom
NIP. 19770330 200501 1 002

Mengesahkan,
Dekan Fakultas Teknik


Ir. Suhartinah, M.T.
NPK. 95 05 246

Mengetahui,
Ketua Program Studi Teknik
Informatika


Yeni Dwi Rahayu, S.T, M.Kom
NPK. 11 03 590



UNIVERSITAS MUHAMMADIYAH JEMBER
FAKULTAS TEKNIK
PROGRAM STUDI INFORMATIKA TERAKREDITASI B
Jl. Karimata 49 Telp/Fax. (0331) 336728 (112)/ 337957 Kotak Pos. 104
JEMBER

PERNYATAAN
ORISINALITAS SKRIPSI

Saya yang bertanda tangan di bawah ini menyatakan bahwa memang benar Skripsi dengan judul Penentuan *Cluster* Optimum untuk Mengelompokkan Kabupaten di Jawa Timur Berdasarkan Tingkat Pengangguran dan Tingkat Kemiskinan Menggunakan Metode *K-Means* merupakan karya orisinal dan bebas dari unsur-unsur PLAGIASI.

Demikian pernyataan ini saya buat dengan sebenarnya. Apabila terbukti terdapat pelanggaran di dalamnya, maka saya bersedia Skripsi ini dibatalkan, Gelar Sarjana Komputer saya dicabut dan saya bersedia menerima sanksi hukum sebagai akibatnya.

Demi untuk kepentingan pengembangan ilmu pengetahuan dan masyarakat, maka saya mengizinkan skripsi ini diterbitkan/dipublikasikan atas sepengetahuan.

Jember, 19 Juli 2019

Mahasiswa,



Nama : Eva Alfariani

NIM : 1510651073

PS : Informatika

Fakultas Teknik

Universitas Muhammadiyah Jember

**PENENTUAN *CLUSTER* OPTIMUM UNTUK MENGELOMPOKAN
KABUPATEN DI JAWA TIMUR BERDASARKAN TINGKAT
PENGANGGURAN DAN TINGKAT KEMISKINAN MENGGUNAKAN
METODE *K-MEANS***

Eva Alfariani

Fakultas Teknik Universitas Muhammadiyah Jember

ABSTRAK

Di Jawa Timur tingkat pengangguran dan tingkat kemiskinan mengalami perubahan setiap tahunnya di beberapa kabupaten/kota. Pengetahuan tentang kelompok wilayah yang terdapat dalam masing-masing kelompok dapat membantu pengambilan kebijakan berbeda-beda setiap tahunnya yang akan dibuat oleh pemerintah provinsi.

Pada penelitian sebelumnya ada pengetahuan untuk mengetahui kelompok wilayah namun menggunakan atribut atau variabel yang berbeda. Penelitian ini menggunakan algoritma *K-Means Clustering* untuk mengetahui kelompok wilayah kabupaten/kota di Jawa Timur dengan pengukuran *performance* menggunakan teknik *Davies-Bouldin Index* untuk mencari *cluster* terbaiknya. Data yang digunakan adalah data tingkat pengangguran dan tingkat kemiskinan kabupaten provinsi Jawa Timur yang memiliki karakteristik 12 fitur dan 456 *record*.

Dari hasil penelitian yang dilakukan diperoleh *cluster* terbaiknya sebesar 3 karena memiliki nilai *Davies-Bouldin Index (DBI)* yang paling minimum yaitu 0.651284382, untuk pengelompokan 2 *cluster* diperoleh nilai *Davies-Bouldin Index (DBI)* 0.732798572, dan pengelompokan 4 *cluster* diperoleh nilai *Davies-Bouldin Index (DBI)* 0.693562132.

Kata kunci: *Data Mining, Clustering, K-Means, Davies-Bouldin Index*

**DETERMINATION OF OPTIMUM CLUSTER TO CATEGORIZE
DISTRICT IN EAST JAVA BASED ON THE LEVEL OF UNEMPLOYMENT
AND POVERTY LEVEL USING K-MEANS**

Eva Alfariani

Faculty of Engineering, University of Muhammadiyah Jember

ABSTRACT

In East Java, unemployment and poverty levels change annually in several districts / cities. Knowledge of the area group included in each group can help policy-making vary annually that will be made by the provincial government.

In a previous study scientific knowledge to know the group of territories but using different attributes or variables. This study uses the K-Means clustering algorithm to determine the groups of districts / cities in East Java performance measurement using techniques Davies-Bouldin Index to find the best cluster. The data used is the data rate of unemployment and poverty rates in East Java provincial district which has a characteristic 12 features and 456 records.

The research results are to be obtained cluster best by 3 because it has the Davies-Bouldin Index (DBI) the minimum is 0.651284382, for grouping 2 cluster values obtained Davies-Bouldin Index (DBI) 0.732798572, and the grouping of four clusters obtained values Davies-Bouldin Index (DBI) 0.693562132.

Keywords: Data Mining, Clustering, K-Means, Davies-Bouldin Index

KATA PENGANTAR

Puji dan syukur kami panjatkan kehadiran Allah SWT yang telah mencurahkan rahmat dan taufik-Nya sehingga penulis dapat menyelesaikan skripsi ini, serta sholawat dan salam semoga selalu tercurah limpahkan kepada Nabi Muhammad SAW.

Skripsi ini berjudul **“Penentuan *Cluster Optimum* untuk Mengelompokkan Kabupaten di Jawa Timur Berdasarkan Tingkat Pengangguran dan Tingkat Kemiskinan Menggunakan Metode *K-Means*”** diajukan sebagai salah satu syarat untuk menyelesaikan pendidikan Studi Strata atau S1 pada Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember.

Penyusunan skripsi ini tidak terlepas dari bantuan dan dorongan dari berbagai pihak yang memungkinkan skripsi ini dapat terselesaikan. Untuk itu rasa terima kasih sedalam-dalamnya penulis haturkan kepada:

1. Dr. Ir. M. Hazmi, D.E.S.S selaku Rektor Universitas Muhammadiyah Jember.
2. Ir. Suhartinah, M.T. selaku Dekan Fakultas Teknik, Universitas Muhammadiyah Jember.
3. Yeni Dwi Rahayu, S.T, M.Kom selaku Ketua Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember.
4. Ari Eko Wardoyo, S.T, M.Kom selaku Dosen Pembimbing Utama dan Agung Nilogiri, S.T, M.Kom selaku Dosen Pembimbing Anggota yang telah banyak memberikan pengarahan, petunjuk, dan membantu selama pelaksanaan penelitian dan penulisan skripsi ini.
5. Terima kasih juga kepada orang tua Ayah dan Ibu yang telah memberikan dukungan moril maupun materil.
6. Teman-teman seperjuangan di Fakultas Teknik Universitas Muhammadiyah Jember.
7. Serta semua pihak yang banyak memberikan bantuan serta motivasi selama melaksanakan penelitian dan penyusunan skripsi.

Penulis menyadari masih banyak sekali kekurangan dan kekeliruan di dalam penulisan skripsi ini, karena penulis hanya manusia biasa yang tak luput dari salah dan dosa. Sehingga penulis secara terbuka menerima segala kritik dan saran yang sangat membangun dan diharapkan untuk perbaikan di masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat sebagai bahan referensi terutama bagi penelitian yang sejenis.

Jember, 19 Juli 2019

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN KEASLIAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
I. PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
II. TINJAUAN PUSTAKA	4
2.1 Jawa Timur	4
2.2 Pengangguran	6
2.3 Kemiskinan.....	7
2.4 <i>Data mining</i>	8
2.5 <i>Clustering</i>	10
2.6 <i>K-Means Clustering</i>	11

2.7	<i>Davies-Bouldin Index</i>	13
2.8	<i>RapidMiner</i>	15
III.	METODE PENELITIAN	16
3.1	Jenis Penelitian	16
3.2	Pengumpulan Data	17
3.3	Proses <i>Clustering</i>	17
3.4	Proses <i>Davies-Bouldin Index</i>	23
3.5	Proses pada <i>RapidMiner</i>	28
3.6	Hasil	28
IV.	HASIL DAN PEMBAHASAN	29
4.1	Implementasi Algoritma <i>K-Means Clustering</i>	30
4.2	Proses Penentuan Jumlah <i>Cluster</i> Optimum	38
4.3	Proses pada <i>RapidMiner</i>	51
4.4	Hasil Pencarian <i>Cluster</i> Menggunakan <i>DBI</i>	56
V.	KESIMPULAN DAN SARAN	58
5.1	Kesimpulan	58
5.2	Saran	59
	DAFTAR PUSTAKA	xiv

DAFTAR GAMBAR

Gambar		Halaman
3.1	Diagram Proses Pencarian <i>Cluster</i> Terbaik	16
3.2	Flowchart Algoritma <i>K-Means Clustering</i>	18
3.3	Tampilan Proses <i>K-Means</i> dan <i>DBI</i>	28
4.1	Tampilan Utama pada <i>RapidMiner</i>	52
4.2	Tampilan Operator <i>Read Excel</i> untuk Menginput Data.....	52
4.3	Proses Menginputkan Data	53
4.4	Tampilan Operator <i>Read Excel</i> yang Sudah Berisi Data.....	53
4.5	Tampilan Mengubah Variabel Bertipe Teks	54
4.6	Tampilan Menambahkan Operator <i>K-Means</i>	54
4.7	Tampilan Menambahkan Operator <i>DBI</i>	55
4.8	Tampilan Mengubah <i>Parameter K-Means</i>	55
4.9	Tampilan Proses <i>K-Means</i> dan <i>DBI</i>	56

DAFTAR TABEL

Tabel		Halaman
2.1	Data LAKIP Dinas Tenaga Kerja Transmigrasi	5
3.1	Data Training	18
3.2	Titik Awal Pusat <i>Centroid</i>	19
3.3	Perhitungan jarak dan pengelompokan data Iterasi Ke-1	20
3.4	<i>Centroid</i> Baru Iterasi ke-1	21
3.5	Perhitungan jarak dan pengelompokan data Iterasi Ke-2	22
3.6	<i>Centroid</i> Baru Iterasi ke-2	22
3.7	Hasil Klasterisasi	22
3.8	Data pada <i>Cluster</i> 1 untuk $k = 3$	23
3.9	Data pada <i>Cluster</i> 2 untuk $k = 3$	23
3.10	Data pada <i>Cluster</i> 3 untuk $k = 3$	23
3.11	<i>Centroid</i> hasil Proses <i>Clustering</i> untuk $k = 3$	24
3.12	<i>SSW</i> untuk $k = 3$	24
3.13	<i>SSB</i> dan <i>Rasio</i> untuk $k = 3$	25
4.1	Data TPT dan TK	29
4.2	Hasil <i>Centroid</i> Pengujian untuk $k = 2$	30
4.3	Hasil Perhitungan Jarak <i>Euclidean</i> untuk $k = 2$	31
4.4	Hasil Anggota <i>Cluster</i> untuk $k = 2$	32
4.5	Hasil <i>Centroid</i> Pengujian untuk $k = 3$	32
4.6	Hasil Perhitungan Jarak <i>Euclidean</i> untuk $k = 3$	33
4.7	Hasil Anggota <i>Cluster</i> untuk $k = 3$	34
4.8	Hasil <i>Centroid</i> Pengujian untuk $k = 4$	35
4.9	Hasil Perhitungan Jarak <i>Euclidean</i> untuk $k = 4$	36
4.10	Hasil Anggota <i>Cluster</i> untuk $k = 4$	37
4.11	Data pada <i>Cluster</i> 1 untuk $k = 2$	38

4.12	Data pada <i>Cluster</i> 2 untuk $k = 2$	39
4.13	<i>Centroid</i> hasil Proses <i>Clustering</i> untuk $k = 2$	40
4.14	<i>SSW</i> untuk $k = 2$	40
4.15	<i>SSB</i> dan <i>Rasio</i> untuk $k = 2$	41
4.16	Data pada <i>Cluster</i> 1 untuk $k = 3$	42
4.17	Data pada <i>Cluster</i> 2 untuk $k = 3$	42
4.18	Data pada <i>Cluster</i> 3 untuk $k = 3$	43
4.19	<i>Centroid</i> hasil Proses <i>Clustering</i> untuk $k = 3$	43
4.20	<i>SSW</i> untuk $k = 3$	43
4.21	<i>SSB</i> dan <i>Rasio</i> untuk $k = 3$	44
4.22	Data pada <i>Cluster</i> 1 untuk $k = 4$	47
4.23	Data pada <i>Cluster</i> 2 untuk $k = 4$	47
4.24	Data pada <i>Cluster</i> 3 untuk $k = 4$	47
4.25	Data pada <i>Cluster</i> 4 untuk $k = 4$	48
4.26	<i>Centroid</i> hasil Proses <i>Clustering</i> untuk $k = 4$	48
4.27	<i>SSW</i> untuk $k = 4$	48
4.28	<i>SSB</i> dan <i>Rasio</i> untuk $k = 4$	49
4.29	Hasil Pengujian untuk $k = 2, k = 3, k = 4$	57