

## PERNYATAAN

Saya menyatakan bahwa skripsi ini bebas plagiat, dan apabila di kemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya bersedia menerima sanksi sesuai ketentuan peraturan perundang-undangan.



Bojonegoro, 13 Agustus 2022



Saniyatus Saidah

NIM :2520180023

# UNUGIRI

## HALAMAN PERSETUJUAN

Nama : Saniyatus Saidah

NIM : 2520180023

Judul : Studi *Clustering* dengan Pendekatan *Data Mining* untuk  
Pengelompokan Sekolah Dasar di Kabupaten Bojonegoro

Telah disetujui dan dinyatakan memenuhi syarat untuk diajukan dalam ujian skripsi.

Bojonegoro, 13 Agustus 2022

Pembimbing I



Denny Nurdiansyah, M.Si.

NIDN. 0726058702

Pembimbing II



Nita Cahyani, S.Si., M.Stat.

NIDN. 0704038906

## HALAMAN PENGESAHAN

Nama : Saniyatus Saidah

NIM : 2520180023

Judul : Studi *Clustering* dengan Pendekatan *Data Mining* untuk  
Pengelompokkan Sekolah Dasar di Kabupaten Bojonegoro

Telah dipertahankan dihadapan penguji pada tanggal 13 September 2022

Dewan Penguji

Ketua



Dr. H. M. Ridlwan Hambali, Lc., MA  
NIDN. 2117056803

Tim Pembimbing

Pembimbing I




Denny Nurdiansyah, S.Si., M.Si.  
NIDN.0726058702

Anggota



Alif Yuanita Kartini, S.Si., M.Si  
NIDN. 0721048606

Pembimbing II



Nita Cahyani, S.Si., M.Stat.  
NIDN. 0704038906

Mengetahui,

Dean Fakultas Sains dan Teknologi



Sunu Wahyudi, M.Pd.  
NIDN.0709058902

Mengetahui,

Ketua Program Studi



Alif Yuanita Kartini, S.Si., M.Si  
NIDN. 0721048606

## MOTTO

Semakin lama hitungan harimu semakin banyak, tapi pada hakikatnya setiap hari berkurang dan semakin mendekati batas akhir, Kehidupan.

(Ibnu Atha'illah)

## PERSEMBAHAN

Ayah dan Ibuku tercinta

**Sholichan (Ayah) dan Salbiyah (Ibu)**

*Terima kasih untuk kasih sayang dan cinta yang telah dan akan selalu diberikan untuk kami. Terima kasih untuk setiap doa dan restu yang selalu dipanjatkan agar kami senantiasa diberikan kemudahan dalam menjalani hidup. Terima kasih untuk setiap usaha yang dilakukan demi memberikan yang terbaik untuk kami. Terima kasih karena telah menjadi ayah dan ibu terbaik bagi kami.*

Adik-Adikku tersayang

**Ahmad Hasib dan Qurrotul Uyun**

*Terima kasih telah hadir dan memberikan warna dalam kehidupan kami. Terima kasih telah menjadi salah satu dari sekian juta alasanmu untuk selalu berusaha memberikan yang terbaik. Terima kasih telah menjadi adik terbaik dalam hidupku.*

# UNUGIRI

## KATA PENGANTAR

Puji syukur kepada tuhan yang Maha Esa atas segala nikmatnya sehingga penulis dapat menyelesaikan studi dan menyusun skripsi yang berjudul “Studi *Clustering* dengan Pendekatan *Data Mining* Untuk Pengelompokan Sekolah Dasar Di Kabupaten Bojonegoro”. Ucapan terima kasih penulis haturkan kepada seluruh pihak yang terlibat dalam penulisan proposal skripsi ini berupa saran, motivasi, dan do’a, yaitu kepada :

1. Bapak K. H. Jauharul Ma’arif, M.Pd.I selaku Rektor Universitas Nahdlatul Ulama Sunan Giri Bojonegoro..
2. Bapak Sunu Wahyudhi, M.Pd selaku Dekan Fakultas Sains dan Teknologi Universitas Nahdlatul Ulama Sunan Giri Bojonegoro.
3. Ibu Alif Yuanita Kartini, M.Si selaku Ketua Program Studi Statistika Universitas Nahdlatul Ulama Sunan Giri Bojonegoro.
4. Bapak Denny Nurdiansyah, M.Si selaku Dosen Pembimbing I dan Ibu Nita Cahyani., M.Stat. selaku Dosen Pembimbing II yang telah memberikan bimbingan ilmu yang bermanfaat serta saran yang membangun demi terselesaikannya skripsi ini.
5. Ibu Nur Mahmudah, M.Stat. selaku Dosen Wali yang telah memberikan bimbingan serta nasehat selama masa perkuliahan.
6. Terkhusus untuk kedua orang tua penulis yang senantiasa memberikan do’a, motivasi dan semangat kepada penulis sejak memulai perkuliahan hingga skripsi ini terselesaikan.

Penyusunan skripsi ini masih jauh dari kesempurnaan, sehingga besar harapan terkait adanya kritik serta saran yang membangun demi perbaikan penulisan skripsi dimasa mendatang. Skripsi ini diharapkan dapat memberikan manfaat kepada berbagai pihak.

Bojonegoro, 13 Agustus 2022

Penulis

Saniyatus Saidah

## ABSTRACT

Saidah, Saniyatus. 2022. Clustering Study with Data Mining Approach for Grouping Elementary Schools in Bojonegoro Regency. Thesis, Department of Statistics, Faculty of Science and Technology, University of Nahdlatul Ulama Sunan Giri. Main Advisor Denny Nurdiansyah, S.Si., M.Sc. and Advisor Nita Cahyani, S.Si., M.Stat. Basic education aims to provide students with basic abilities to develop their lives as individuals, community members, citizens, and members of humanity and prepare students to attend education. The implementation of national education must ensure equity and improve the quality of education. However, based on regional education balance data (NDP) in 2021, the capacity and facilities of primary school education have not been evenly distributed and to anticipate this, it is necessary to group elementary schools based on capacity and facilities. The purpose of this study is to compare clustering methods to obtain the best method so that it can be applied to classify primary schools based on capacity and school facilities. This study applies 3 clustering methods, the first is K-Means, K-medoid, and Random Clustering which will be compared to get the best method. The data source used is secondary data consisting of the number of students, teachers, classrooms, and groups taken from the Ministry of Research, Technology and Higher Education's dapodik website in 2020. The results of the comparison of methods that have been carried out to cluster elementary school data, the best method is the K-Means method. with the lowest value of the performance vector - 3569.328. formed 5 clusters including cluster\_0 with 177 elementary school members, cluster\_1 with 310 elementary schools, cluster\_2 with 236 elementary school members, cluster\_3 with 14 elementary school members, cluster\_4 with 176 elementary school members. The conclusion of this study is that for equal distribution of capacity and facilities, elementary schools in Bojonegoro are clustered into 5 cluster groups by applying the K-Means method as the best method.

Keywords: Elementary School, K-Means, K-Mendoid, Random Clustering



## ABSTRAK

Saidah, Saniyatus. 2022. *Studi Clustering dengan Pendekatan Data Mining untuk Pengelompokan Sekolah Dasar di Kabupaten Bojonegoro*. Skripsi, Jurusan Statistika Fakultas Sains dan Teknologi Universitas Nahdlatul Ulama Sunan Giri. Pembimbing Utama Denny Nurdiansyah, S.Si., M.Si. dan Pembimbing Pendamping Nita Cahyani, S.Si., M.Stat. Pendidikan dasar bertujuan untuk memberikan bekal kemampuan dasar kepada peserta didik untuk mengembangkan kehidupannya sebagai pribadi, anggota masyarakat, warga negara, dan anggota umat manusia serta mempersiapkan peserta didik untuk mengikuti pendidikan. Pelaksanaan pendidikan nasional harus menjamin pemerataan dan peningkatan mutu pendidikan. Namun berdasarkan data neraca pendidikan daerah (NDP) pada tahun 2020 kapasitas dan fasilitas pendidikan sekolah dasar belum merata secara menyeluruh dan untuk mengantisipasi hal tersebut diperlukan pengelompokan sekolah dasar berdasarkan kapasitas dan fasilitas. Tujuan penelitian ini adalah membandingkan metode-metode clustering untuk mendapatkan metode terbaik sehingga bisa diterapkan untuk mengelompokkan sekolah dasar berdasarkan kapasitas dan fasilitas sekolah. Penelitian ini menerapkan 3 metode clustering yang pertama K-Means, K-medoid, dan Random Clustering yang akan diperbandingkan untuk mendapatkan metode terbaik. Sumber data yang digunakan adalah data sekunder yang terdiri dari jumlah peserta didik, guru, ruang kelas, dan rombel yang diambil dari web dapodik kemenristekdikti tahun 2021. Hasil perbandingan metode yang telah dilakukan untuk mengkluster data sekolah dasar, metode yang terbaik adalah metode K-Means dengan nilai dari *performance vector* paling rendah -3569.328. membentuk 5 *cluster* diantaranya cluster\_0 dengan beranggotakan 177 sekolah dasar, cluster\_1 yang beranggotakan 310 sekolah dasar, cluster\_2 dengan jumlah anggota 236 sekolah dasar, cluster\_3 dengan jumlah anggota 14 sekolah dasar, cluster\_4 dengan jumlah anggota 176 sekolah dasar. Kesimpulan dari penelitian ini adalah untuk pemerataan kapasitas dan fasilitas SD di Bojonegoro diklusterkan menjadi 5 kelompok cluster dengan menerapkan metode K-Means sebagai metode terbaik.

Kata Kunci: Sekolah Dasar, K-Means, K-Mendoid, Random Clustering

# DAFTAR ISI

	Halaman
<b>HALAMAN SAMPUL LUAR (COVER)</b> .....	<b>i</b>
<b>HALAMAN SAMPUL DALAM</b> .....	<b>ii</b>
<b>PERNYATAAN</b> .....	<b>iii</b>
<b>HALAMAN PERSETUJUAN</b> .....	<b>iv</b>
<b>HALAMAN PENGESAHAN</b> .....	<b>v</b>
<b>MOTTO DAN PERSEMBAHAN</b> .....	<b>vi</b>
<b>KATA PENGANTAR</b> .....	<b>vii</b>
<b>ABSTRACT</b> .....	<b>viii</b>
<b>ABSTRAK</b> .....	<b>ix</b>
<b>DAFTAR ISI</b> .....	<b>x</b>
<b>DAFTAR TABEL</b> .....	<b>xiii</b>
<b>DAFTAR GAMBAR</b> .....	<b>xiv</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xv</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	7
1.3 Tujuan Penelitian .....	7
1.4 Manfaat Penelitian .....	7
<b>BAB II TINJAUAN PUSTAKA DAN DASAR TEORI</b> .....	<b>9</b>
2.1 Tinjauan Pustaka .....	9
2.2 Dasar Teori .....	11
2.2.1 Guru.....	11
2.2.2 Peserta Didik .....	11
2.2.3 Ruang Kelas .....	12
2.2.4 Rombel (Rombongan Belajar) .....	12
2.2.5 <i>Data Mining</i> .....	12
2.2.6 Metode Clustering .....	13
2.2.7 Metode K-Means.....	14
2.2.8 Metode K-Medoid .....	16
2.2.9 <i>Random Clustering</i> .....	19



2.2.10 <i>Performance Vector</i> .....	19
2.2.11 Klasifikasi Pengelompokkan <i>Clustering</i> .....	20
2.2.12 <i>Software</i> RapidMiner Studio .....	21
<b>BAB III METODOLOGI PENELITIAN</b> .....	<b>25</b>
3.1 Sumber Data.....	25
3.2 Variabel Penelitian .....	25
3.3 Langkah Analisis.....	25
3.4 Diagram Alir .....	36
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	<b>38</b>
4.1 Statistika Deskriptif Data Pengamatan.....	38
4.2 Pengelompoksn SD di Kabupaten Bojonegoro .....	42
4.2.1 Metode K-Means .....	42
4.2.2 Metode K-Medoid .....	45
4.2.3 Metode <i>Random Clustering</i> .....	49
4.3 Hasil Perbandingan Metode <i>Clustering</i> .....	53
<b>BAB V KESIMPULAN DAN SARAN</b> .....	<b>49</b>
5.1. Kesimpulan .....	49
5.2. Saran.....	60
<b>DAFTAR PUSTAKA</b> .....	<b>61</b>
<b>LAMPIRAN</b> .....	<b>67</b>



**UNUGIRI**

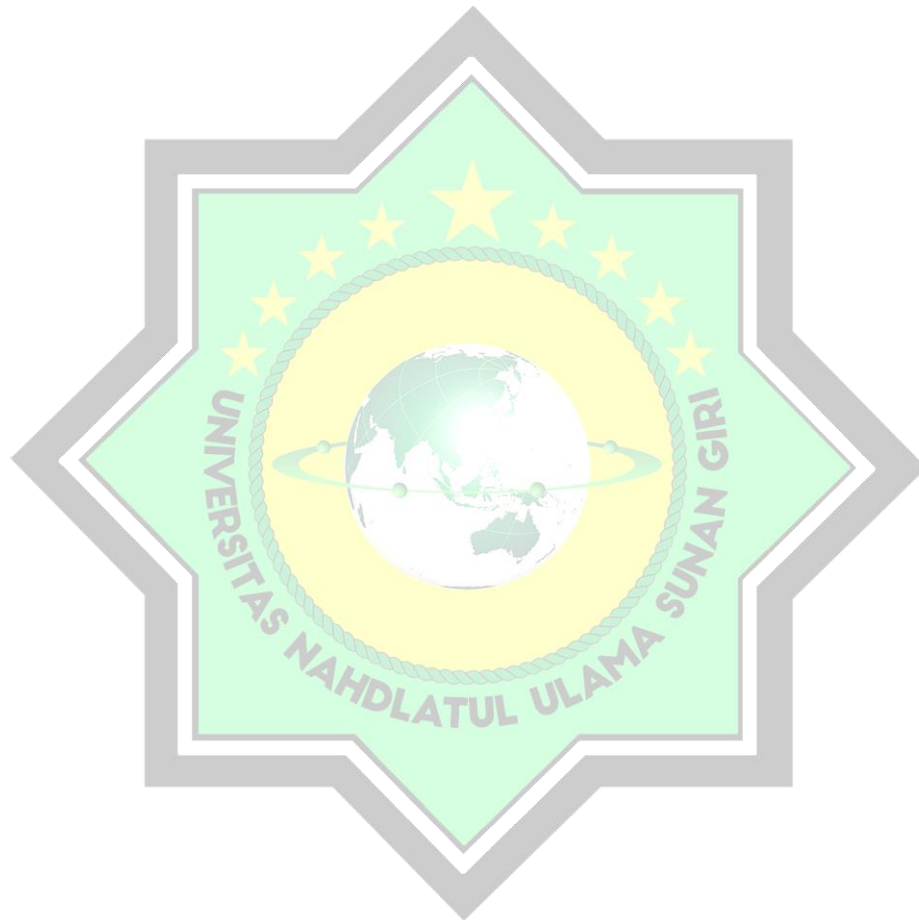
## DAFTAR TABEL

Tabel		Halaman
2.1	Penelitian Terdahulu.....	9
2.2	Kriteria Klasifikasi Model <i>Clustering</i> Terbaik.....	21
3.1	Variabel Penelitian.....	25
4.1	Sampel Data Penelitian <i>Clustering</i> .....	38
4.2	Deskriptif Statistik .....	40
4.3	Zscore Deskriptif Statistik.....	41
4.4	Studi Penyelidikan Metode K-Means.....	43
4.5	Hasil Pengelompokan SD di Bojonegoro dengan Metode K-Means pada $k=5$ .....	44
4.6	Studi Penyelidikan Metode K-Medoid.....	47
4.7	Hasil Pengelompokan SD di Bojonegoro dengan Metode K-Medoid pada $k=5$ .....	47
4.8	Studi Penyelidikan Metode <i>Random Clustering</i> .....	51
4.9	Hasil Pengelompokan SD di Bojonegoro dengan Metode <i>Random Clustering</i> pada $k=5$ .....	51
4.10	Perbandingan Metode-Metode <i>Clustering</i> .....	53
4.11	Deskriptif <i>Cluster</i> dari <i>Clustering</i> K-Means.....	54
4.12	Anggota kelompok <i>cluster</i> SD menggunakan metode K-Means.....	56

## DAFTAR GAMBAR

Gambar	Halaman
2.1 <i>Output</i> Metode Clustering.....	14
2.2 Metode <i>Clustering</i> K-Means.....	15
2.3 <i>Output</i> Metode <i>Clustering</i> K-Medoid.....	18
2.4 Sudut $\theta_e$ adalah sudut yang termuat dalam $e$ dan sudut $e \in E..$	19
2.5 Tampilan Menu Utama RapidMiner.....	23
2.6 Platform RapidMiner Untuk Penerapan <i>Data Mining</i> Pada <i>Segmentation</i>	24
3.1 Tampilan Awal Software RapidMiner.....	28
3.2 Tampilan Menu Utama Rapidminer.....	28
3.3 Import Data <i>Clustering</i> .....	29
3.4 Pemilihan <i>Role</i> Data.....	30
3.5 <i>Retrieve</i> Data <i>Clustering</i> .....	30
3.6 Tampilan Menu <i>Multiply</i> Data <i>Clustering</i> .....	31
3.7 Tampilan <i>Design</i> Proses <i>Data To Similarity</i> .....	32
3.8 Tampilan <i>Design</i> Proses Utama Metode <i>Clustering</i> .....	33
3.9 <i>Design</i> Proses K-Means <i>Clustering</i> .....	33
3.10 <i>Design</i> Proses K-Medoid <i>Clustering</i> .....	33
3.11 <i>Design</i> Proses <i>Random Clustering</i> .....	34
3.12 <i>Design</i> Proses <i>Performance Vector</i> .....	34
3.13 Diagram Alir Proses Penelitian.....	36
4.1 Boxplot.....	40
4.2 <i>Output Cluster</i> Model K-Means.....	41
4.3 <i>Output Performance Vector</i> Metode K-Means.....	42
4.4 <i>Output Cluster</i> Model K-Medoid .....	44
4.5 <i>Output Scatter Clustering</i> K-Medoid.....	45
4.6 <i>Output Performance Vector</i> Metode K-Medoid.....	45
4.7 <i>Output Cluster</i> Model <i>Random Clustering</i> .....	48

4.8	<i>Output Scatter Clustering Random Clustering.....</i>	49
4.9	<i>Output Performance Vector Random Clustering.....</i>	49
4.10	<i>Perbandingan Metode Clustering.....</i>	52



**UNUGIRI**

## DAFTAR LAMPIRAN

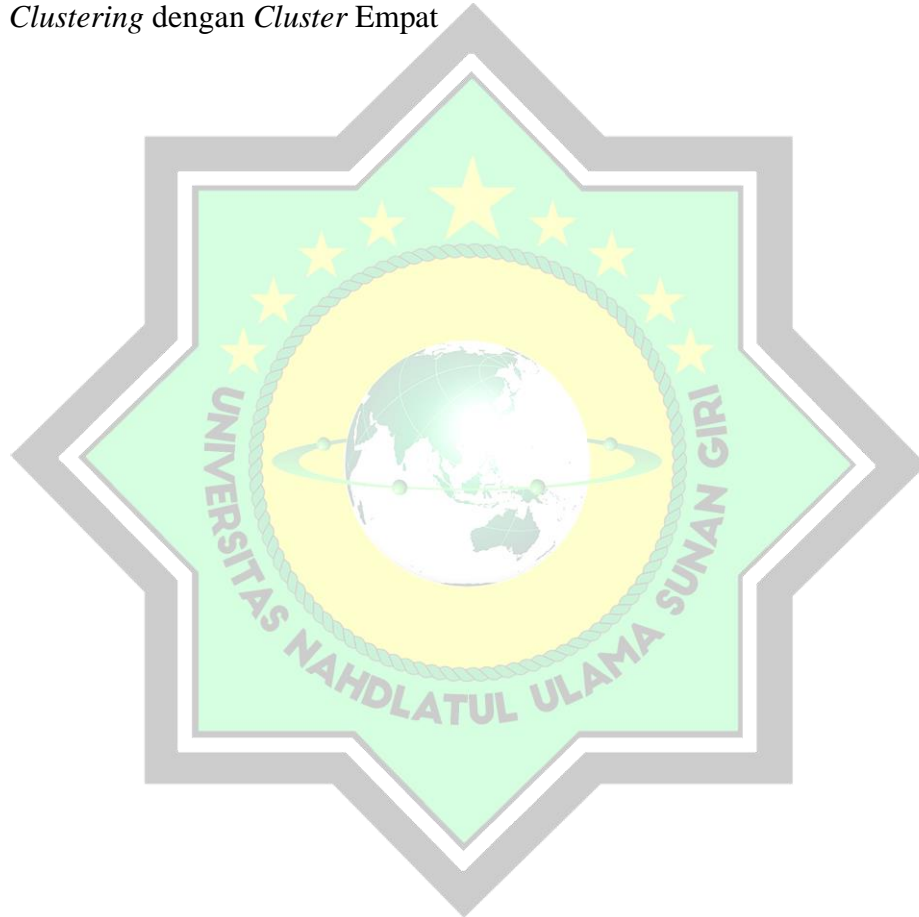
### Lampiran

- 1 5% dari Data Pengamatan SD Kabupaten Bojonegoro
- 2 Design Process Software Rapidminer untuk Metode K-Means dengan *Cluster* Dua
- 3 *Output ExampleSet* Software Rapidminer untuk Metode K-Means dengan *Cluster* Dua
- 4 *Output ExampleSet* Software Rapidminer untuk Metode K-Means dengan *Cluster* Dua
- 5 *Output Performance Vector* Software Rapidminer untuk Metode K-Means dengan *Cluster* Dua
- 6 Design Process Software Rapidminer untuk Metode K-Means dengan *Cluster* Tiga
- 7 *Output ExampleSet* Software Rapidminer untuk Metode K-Means dengan *Cluster* Tiga
- 8 *Output ExampleSet* Software Rapidminer untuk Metode K-Means dengan *Cluster* Tiga
- 9 *Output Performance Vector* Software Rapidminer untuk Metode K-Means dengan *Cluster* Tiga
- 10 Design Process Software Rapidminer untuk Metode K-Means dengan *Cluster* Empat
- 11 *Output ExampleSet* Software Rapidminer untuk Metode K-Means dengan *Cluster* Empat
- 12 *Output ExampleSet* Software Rapidminer untuk Metode K-Means dengan *Cluster* Empat
- 13 *Output Performance Vector* Software Rapidminer untuk Metode K-Means dengan *Cluster* Empat
- 14 *Design Process* Software Rapidminer untuk Metode K-Medoid dengan *Cluster* Dua
- 15 *Output ExampleSet* Software Rapidminer untuk Metode K-Medoid dengan *Cluster* Dua
- 16 *Output ExampleSet* Software Rapidminer untuk Metode K-Medoid dengan *Cluster* Dua

- 17 *Output Performance Vector* Software Rapidminer untuk Metode K-Medoid dengan *Cluster* Dua
- 18 *Design Proses* Software Rapidminer untuk Metode K-Medoid dengan *Cluster* Tiga
- 19 *Output ExampleSet* Software Rapidminer untuk Metode K-Medoid dengan *Cluster* Tiga
- 20 *Output ExampleSet* Software Rapidminer untuk Metode K-Medoid dengan *Cluster* Tiga
- 21 *Output Performance Vector* Software Rapidminer untuk Metode K-Medoid dengan *Cluster* Tiga
- 22 *Design Proses* Software Rapidminer untuk Metode K-Medoid dengan *Cluster* Empat
- 23 *Output ExampleSet* Software Rapidminer untuk Metode K-Medoid dengan *Cluster* Empat
- 24 *Output ExampleSet* Software Rapidminer untuk Metode K-Medoid dengan *Cluster* Empat
- 25 *Output Performance Vector* Software Rapidminer untuk Metode K-Medoid dengan *Cluster* Empat
- 26 *Design Proses* Software Rapidminer untuk Metode *Random Clustering* dengan *Cluster* Dua
- 27 *Output ExampleSet* Software Rapidminer untuk Metode *Random Clustering* dengan *Cluster* Dua
- 28 *Output ExampleSet* Software Rapidminer untuk Metode *Random Clustering* dengan *Cluster* Dua
- 29 *Output Performance Vector* Software Rapidminer untuk Metode *Random Clustering* dengan *Cluster* Dua
- 30 *Design Process* Software Rapidminer untuk Metode *Random Clustering* dengan *Cluster* Tiga
- 31 *Output ExampleSet* Software Rapidminer untuk Metode *Random Clustering* dengan *Cluster* Tiga
- 32 *Output ExampleSet* Software Rapidminer untuk Metode *Random Clustering* dengan *Cluster* Tiga
- 33 *Output Performance Vector* Software Rapidminer untuk Metode *Random Clustering* dengan *Cluster* Tiga



- 34 *Design Process Software Rapidminer untuk Metode Random Clustering dengan Cluster Empat*
- 35 *Output ExampleSet Software Rapidminer untuk Metode Random Clustering dengan Cluster Empat*
- 36 *Output ExampleSet Software Rapidminer untuk Metode Random Clustering dengan Cluster Empat*
- 37 *Output Performance Vector Software Rapidminer untuk Metode Random Clustering dengan Cluster Empat*



**UNUGIRI**